A RAND NOTE

Estimating Posthospital Use for Principal and Secondary Diagnoses

Joanna Zorn Heilbrunn, Neal Thomas

August 1990

RAND

The research described in this report was sponsored by the University of Minnesota under Subcontract 17-C-98891/5-02 awarded by the Health Care Finanacing Administration, U.S. Department of Health and Human Services.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

A RAND NOTE

N-3002-MN

Estimating Posthospital Use for Principal and Secondary Diagnoses

Joanna Zorn Hellbrunn, Neal Thomas

August 1990

Prepared for the
University of Minnesota

RAND

PREFACE

This Note documents the relationship between the principal and secondary diagnoses of Medicare hospital patients and the propensity of these patients to use home health agencies, skilled nursing facilities, and rehabilitation facilities after discharge from the hospital. The findings are directed to policymakers considering payment systems that bundle posthospital services with hospital payments, especially systems that permit voluntary participation by hospitals and patients. A possible use for the results of this research is to help researchers identify a fairly small set of common secondary diagnoses that appear to be related to increased (or decreased) use of different types of posthospital care among those sampled in quality of care studies. More careful study of patients with these diagnoses could provide a first step toward the creation of more sensitive measures of the access to and quality of posthospital care.

This Note is part of a research project on the "natural history" of episodes of care among Medicare beneficiaries being conducted at The RAND Corporation and the University of Minnesota. The authors thank Lisa Rubenstein for her helpful comments and consulting. Daniel Relles reviewed this Note and offered many helpful comments. His suggestions regarding the demonstration of the predictive capability of the secondary diagnoses were especially helpful.

The work reported in this Note was supported by Health Care Financing Administration cooperative agreement 17-C-98891/5-02.

SUMMARY

One concern with any prospective payment system that bundles payment of posthospital care with primary hospital care is the possibility that hospital administrators and doctors will learn to predict which types of patients will require posthospital care. Since any extra posthospital care would have to be paid by the hospital out of the fixed prospective payment, predictable variation in the use of posthospital care within payment categories might create an incentive to exclude patients likely to require the most care. This Note demonstrates that knowledge of a patient's secondary diagnoses provides additional information about the patient's use of posthospital care beyond that contained in the principal diagnosis. Unless a small number of secondary diagnoses can be identified that account for most of this additional variability in posthospital use, refinement of the current diagnostic related group (DRG) system to account for differences among patients with differing secondary diagnoses will not be feasible. The study provides concrete evidence that knowledgeable administrators can skim desirable patients, not only from the current hospital DRG categories but also from categories defined by the principal diagnoses contained in the current DRG categories.

The results presented are crude, providing a very conservative lower bound on the success rate of a knowledgeable predictor. The crudeness is due both to the type and quality of data available for this investigation and to the necessary simplifications imposed by statistical modeling of such a complex process. The only additional patient characteristics used to further distinguish between patients with the same DRG and principal diagnosis are the secondary diagnoses recorded on the Medicare MEDPAR file. There is evidence that the secondary diagnoses are not reliably recorded, which, if true, should further reduce the predictive capability of the secondary codes we used. Patterns of use among patients with different secondary codes are generally consistent with clinical experience providing evidence that the correlation between certain secondary diagnoses and an increase in the use of posthospital care is not an artifact of the haphazard recording of secondary diagnoses. Secondary diagnoses represent a small portion of the relevant information available on a patient at the time of admission. Other characteristics, such as the recent use of a nursing facility or home health service, are expected to be much better predictors of the use of posthospital care.

The analyses were performed separately for the five DRG categories with highest posthospital use: DRG 14, stroke; DRG 88, chronic obstructive pulmonary disease;

DRG 127, heart failure and shock; DRG 209, major joint and limb reattachment procedures; and DRG 210, other hip and femur procedures to patients age 70 or older. Regression methods were used to assess whether the secondary diagnoses contain additional information about posthospital use not measured in the principal diagnoses. Models were fit using only the principal diagnoses, and then the secondary diagnoses were added to these models to determine whether they provided additional information about posthospital use. Formal statistical analyses give strong evidence against the hypothesis that the secondary diagnoses are unrelated to posthospital use once differences related to principal diagnoses are eliminated. As measured by these significance tests, and the size of the parameter estimates, the secondary diagnoses are most useful for predicting skilled nursing facility use. Given the relatively large samples available for this study, rejection of this naive hypothesis is not surprising. Judging by the size of their estimates and standard errors, most secondary diagnoses do not appear to be related to the use of posthospital care, but in each DRG approximately five to ten secondaries are predictive of the different types of posthospital care. (The presence of a secondary diagnosis can raise the odds of one type of posthospital use while lowering the odds of another type of service.) Many of these secondary diagnoses, such as diabetes, appear frequently in all of the DRG categories, and their presence typically has a consistent effect on the use of posthospital care across the DRGs. Differences in the parameter estimates for the principal and secondary diagnoses are roughly equal in magnitude.

To indicate the amount of differentiation among patients that is possible using primary and secondary diagnoses, models were fit to a randomly selected half of the data and then used to predict the posthospital use in the other half of the data. The range of differing rates of posthospital use between patients with high and low predicted probabilities of use is large and should still serve only as a lower bound for the predictive capability of a knowledgeable investigator or administrator with access to more and better quality data.

These findings are directed to policymakers considering bundled payment systems, especially systems that permit voluntary participation. Another possible use for the results in this Note is to help researchers identify a small set of common secondary diagnoses that appear to be related to increased (or decreased) use of different types of posthospital care. More careful study of patients with these diagnosed conditions could provide a first step toward the creation of more sensitive measures of the access to and quality of posthospital care.

CONTENTS

FIGURES

TABLES

I. INTRODUCTION

The purpose of this Note is to demonstrate that the secondary diagnoses routinely reported to the Health Care Financing Administration (HCFA) for Medicare patient hospital admissions contain additional predictive information about these patients' use of posthospital services beyond that contained in the DRG classifications and principal diagnoses. The analysis is based on data contained in a 20 percent sample of all Medicare hospital discharges that occurred between July 1, 1984, and June 30, 1985. This file is known as the MEDPAR file. It has been linked by researchers at RAND to several other data sources containing information on posthospital use by these patients. The extended definition of an episode of care (Neu and Harrison, 1986) was used to determine whether a patient's use of Home Health Agency (HHA), skilled nursing facility (SNF), or rehabilitation hospital (REHAB) care qualifies for inclusion as posthospital use. For each hospital admission in the file, the following classification variables are available: DRG, principal diagnosis, and up to four secondary diagnoses. Other potential predictors, such as age and sex, are also contained in the MEDPAR file, but they were not used in this analysis. Indicators of posthospital care provided by a SNF, HHA, or REHAB are also available for each patient.

A complete analysis is given for patients in DRG 014, stroke victims. This DRG is used for the principal presentation because these patients account for a substantial portion of all posthospital care use. Brief summaries of similar analyses for several other DRG categories are presented in Appendix A. A terse description of the principal and secondary diagnoses for DRG 014 is given in Tables 1.1 and 1.2. The percentages of DRG 14 patients who use HHA, SNF, and REHAB care are also shown in these tables. These rates are given for all DRG 14 patients, for patients grouped by principal diagnosis, and for patients grouped by selected secondary diagnoses. The utilization rates vary substantially across the secondary diagnoses. Analyses, not elaborated on in this Note, show that these differences cannot be explained by simple models of variation resulting from sampling. The fact that the posthospital utilization rates appear to vary more among secondary diagnoses than among principal diagnoses based on the range of the observed rates should not be taken to imply that the secondary diagnoses will be better predictors of posthospital care. Even if posthospital care use were completely unrelated to principal and secondary diagnoses, this phenomenon would still be expected; the greater number of

secondary diagnoses increases the expected range of posthospital rates caused by sampling fluctuations.

The technical sections that follow address the issue of whether the difference in utilization rates among secondary diagnoses can be explained by their association with some particular principal diagnoses, or whether they have their own distinct predictive ability.

Table 1.1
PRINCIPAL DIAGNOSES

ICD9-CM	Description	Live Discharge	Rehab. Use, %	SNF Use, %	HHA Use, %	No Use, %
430	Subarachnoid hemorrhage	428	2.6	12.4	13.8	70.2
431	Intracerebral hemorrhage	2,784	8.5	19.4	18.9	53.2
4320	Nontraum. extradural hem.	15	0.0	6.7	0.0	93.3
4321	Subdural hemorrhage	332	2.1	7.2	9.3	81.3
4329	Intracranial hem. nos.	90	4.4	8.9	18.9	67.8
4340	Cerebral thrombosis	6,856	6.9	15.5	23.2	54.3
4341	Cerebral embolism	3,257	7.9	11.7	22.5	57.9
4349	Cerebral artery occlus. nos.	14,842	8.8	14.7	22.8	53.7
436	CVA	22,211	5.6	13.2	22.4	58.9
4373	Nonrupt. cereb. aneurysm	179	1.1	4.5	7.8	86.6
7843	Aphasia	161	3.7	9.9	15.5	70.8
Total		51,155	6.9	14.1	22.2	56.8

The analysis for DRG 014 is divided into two sections. In the first section, the relationship between principal and secondary diagnoses is examined. Although these diagnoses are dependent, the dependence is not very strong. Each secondary diagnosis does not occur with only one or two principal diagnoses. The differences in rates of posthospital care use between patients with different secondary diagnoses cannot be fully explained by differences in the principal diagnosis with which they most often occur. There is also some weak dependence among the different secondary diagnoses. This dependence arises because up to four different secondary diagnoses can be recorded for a patient, and someone with any secondary diagnosis recorded is more likely to have some other secondary diagnosis as well. Patients commonly have two or three secondary diagnoses, but no pair of secondary diagnoses occurs together with high concordance.

The second section directly verifies our claim that the secondary diagnoses contain additional information about use of posthospital care beyond that contained in the principal diagnosis. This was done by fitting logistic regression models that include

Table 1.2

SECONDARY DIAGNOSES

Code	Description	Live Discharge	Rehab. Use, %	SNF Use, %	HHA Use, %	No Use, %
V10	History of mlg. neoplasm	793	5.8	15.4	23.2	55.6
V45	Other postsurgical states (usually pacemaker)	894	8.4	12.2	20.6	58.8
41	Bacterial infection	1,962	6.1	26.7	21.9	45.4
244	Acquired hypothyroidism	566	8.0	16.8	21.6	53.7
250	Diabetes mellitus	8,416	7.9	14.2	25.2	52.7
276	Disord.—fluid, acid balance	3,353	4.3	19.0	21.0	55.8
278	Obesity	512	9.4	11.7	21.5	57.4
285	Unspecified anemias	664	3.2	17.5	24.8	54.5
290	Senile & presenile—organic	933	1.9	19.2	22.3	56.6
298	Nonorganic psychoses	565	2.8	15.4	19.8	61.9
310	Nonpsych. mental—organic	1,298	2.1	16.9	20.1	60.9
331	Cerebral degeneration	1,218	3.3	14.9	22.2	59.6
332	Parkinson's disease	601	4.3	18.6	22.6	54.4
342	Hemiplegia	10,792	13.7	19.8	22.2	44.3
344	Other paralytic syndromes	978	7.9	16.6	24.8	50.7
345	Epilepsy	548	5.1	14.4	23.7	56.8
401	Essential hypertension	12,458	9.1	12.8	23.0	55.0
402	Hypertensive heart disease	2,610	8.4	14.1	25.8	51.7
410	Acute myocardial infarction	566	11.1	14.1	23.1	51.6
412	Old myocardial infarction	964	9.3	10.4	20.6	59.6
413	Angina pectoris	987	8.0	10.2	21.0	60.8
414	Other chron. isch. heart dis. (coronary atherosclerosis)	6,876	6.4	13.6	21.9	58.1
424	Endocardium—other diseases	897	8.8	13.3	21.1	56.9
426	Conduction disorders	1,129	8.1	12.3	25.9	53.8
427	Cardiac dysrythmias	7,843	7.9	16.3	21.3	54.5
428	Heart failure	3,745	5.9	18.3	23.3	52.5
429	Heart disease—ill-defined	2,419	5.7	14.6	22.0	57.8
433	Precerebral art. occ. & sten.	1,959	8.5	8.1	21.3	62.1
434	Occlusion of cerebral artery	793	7.3	14.1	24.8	53.7
435	Transient cerebral ischem.	1,145	2.4	6.4	20.2	71.1
436	Acute cerebrovascular dis.	705	6.7	16.7	23.7	52.9
437	Other cerebrovascular dis.	2,511	4.8	12.4	23.4	59.3
438	Late effects of cerebral dis.	1,405	6.5	16.3	23.3	53.8
440	Atherosclerosis	2,021	5.3	12.9	22.5	59.3
443	Peripheral vascular disease	566	9.5	13.1	23.7	53.7
486	Pneumonia, organism unspec.	1,128	4.6	27.3	19.0	49.1
496	Chronic airway obstruction	2,254	6.6	14.2	22.3	56.9
507	Pneumonitis—solids & liquids	642	7.2	32.6	14.3	46.0
599	Disord. urinary tr. & urethra	5,212	7.1	23.2	23.0	46.7
715	Osteoartrosis	1,539	6.5	14.4	21.6	57.6
780	General symptoms (coma & convulsions, etc.)	3,059	3.8	17.1	19.5	59.6
784	Sympt. involv head & neck (aphasia, speech disturb.)	4,295	12.4	18.6	21.3	47.7
787	Sympt. involv. digestive sys. (usually swallowing difficulty)	673	9.8	25.1	21.7	43.4
788	Sympt. involv. urinary system	620	6.8	21.1	26.5	45.6
None	No secondary diagnoses	4,676	4.2	9.2	21.7	64.9

terms for the principal diagnosis and the secondary diagnoses to the posthospital care utilization rates. A formal statistical test of the null hypothesis (that the secondary diagnoses have no additional explanatory power) was obtained bv comparing the fit of

these models to that of models that include only the principal diagnoses. These tests are very significant. The parameter estimates for several of the secondary diagnoses indicate that their presence substantially changes the odds that a patient will use posthospital care. This was confirmed by using the estimated models to predict posthospital care for a large sample of patients excluded while the models were fit. Because the secondary diagnoses are a poor substitute for the information available to decisionmakers at hospitals, these models are only a very crude attempt to use the information contained in secondary diagnoses to indicate the potential gains in predictive capability available to an administrator with much superior information. The results are intended only as a feasibility study for someone interested in predicting use of posthospital care using patient characteristics.

II. CORRELATION AMONG DIAGNOSES

This section analyzes the relationship between principal and secondary diagnoses. If each secondary diagnosis occurs with only one or two principal diagnoses, then the secondary diagnoses, as well as the observed differences in use rates for different secondary diagnoses, could be predicted from the principal diagnosis alone. The results of this section show that despite some dependence between the diagnoses, no small group of diagnoses exists that is highly predictive of the presence of the remaining diagnoses.

Since each patient must have a principal diagnosis, the principal diagnosis can be represented by n_d 0/1 indicator variables, $D_1,...,D_{nd}$, where n_d is the number of prominent principal diagnoses included in Table 1.1. All other principal diagnoses occurring in the DRG are represented by setting each $D_1,...,D_{n_d}$ equal to zero. For some DRGs, including DRG 14, the other category is very small, so one of the principal diagnoses was used as the reference so that standard errors would not be misleadingly inflated. A patient may have several secondary diagnoses, so each of the n_s secondary diagnoses will be represented by 0/1 indicator variables, $S_1,...,S_{nd}$. In addition to the secondary diagnoses listed in Table 1.2, an extra category was created to represent the presence of a secondary diagnosis not included in the list of prominent secondaries. The value of n_s is thus the number of secondary diagnoses in Table 1.2 plus one. Up to four of the S_i may be one, indicating that the patient had four secondary diagnoses. Unlike the principal diagnosis, all of the S_i may also be zero (in fact, this happens often), indicating that the patient had no recorded secondary diagnoses. For DRG 14, $n_d = 11$, and $n_s = 45$.

If each secondary diagnosis is independent of the principal diagnosis, then the additional predictive capability of the secondary diagnoses is the same as their predictive capability excluding the principal diagnosis. For posthospital use data, this means that the large differences in posthospital use for patients with the different secondary diagnoses seen in Table 1.2 cannot be explained by differences in the patients' principal diagnoses. Since the principal diagnosis is a discrete variable with no natural numerical scale and many observations at each value, the conditional distribution of each secondary diagnosis given the principal diagnosis was estimated by the sample mean (proportion for 0/1 data) for each different principal diagnosis.

Table 2.1 contains the estimated distributions for each secondary diagnosis conditional on the principal diagnosis. Rows correspond to the presence of a particular secondary diagnosis, and columns correspond to the different principal diagnoses. Each

entry is the proportion of patients with the secondary diagnosis from among those patients with the principal diagnosis given by the column. The conditional distribution of each secondary diagnosis given the principal diagnoses is given by an entire row in the table. Although the sample sizes given in Table 1.2 are large, sampling variation can still produce differences between the estimated proportions. This is particularly true because for many of the secondary diagnoses, these proportions, and potentially important differences between them, are very small. To aid in the search for secondary diagnoses that are strongly related to the principal diagnosis, a chi square statistic was computed for each estimated conditional distribution. These statistics are presented in the final column of Table 2.1. Under the hypothesis of independence, each secondary diagnosis has the same probability of appearing given any particular principal diagnosis, and this chi square statistic should have approximately 11 degrees of freedom. A large value for the test statistic under this assumption is 20.

Examination of these chi square statistics in Table 2.1 conclusively rules out the independence assumption. Since the proportions are very small, there is some doubt about the accuracy of the chi square approximation, but a small number of simulations using the proportions and the sample sizes from Table 2.1 showed that the larger chi square statistics do not result from a poor approximation to the null distribution of the test statistic. The secondary diagnosis showing the greatest dependence on the principal diagnosis as measured by the chi square statistic is 427, cardiac dysrythmias. The chi square statistic is large because 48 percent of the patients with cerebral embolism (4341) listed as their principal diagnosis have a secondary diagnosis of cardiac dysrythmias, whereas fewer than 10 percent of patients with principal diagnoses of 4320, 4321, 4373, and 7843 have secondary diagnosis 427 listed. The principal diagnosis with which the secondary diagnosis 427 occurs most commonly has higher than average posthospital use; the principal diagnoses with which it rarely occurs have lower than average posthospital use. The higher posthospital use by patients with secondary diagnosis 427 can thus be largely explained by differences in principal diagnosis. Similar patterns can be found for a few of the other secondary diagnoses with large chi square statistics, but despite the unequivocal evidence that the principal and secondary diagnoses are dependent, it is not clear whether the dependency is strong enough to explain all of the variation in posthospital use among the secondary diagnoses in Table 1.2. This judgment cannot be readily drawn from the examination of Table 2.1 alone, but the more direct methods presented in Sec. III show that this dependence between secondary and principal

Table 2.1

DISTRIBUTION OF SECONDARY DIAGNOSIS CONDITIONAL ON PRINCIPAL DIAGNOSIS

Code	430	431	4320	4321	4329	4340	4341	4349	436	4373	7843	Chi Square
10	.011	.015	.118	.015	.012	.017	.010	.018	.015	.011	.025	28.3
45	.006	.013	.029	.033	.012	.014	.034	.017	.015	.016	.012	89.5
41	.037	.039	.029	.021	.023	.038	.026	.040	.036	.016	.006	28.5
244	.012	.008	.029	.008	.003	.011	.009	.012	.009	.011	.012	18.0
250	.069	.106	.118	.077	.064	.183	.133	.172	.162	.053	.153	272.9
276	.072	.055	.059	.067	.052	.061	.043	.067	.074	.079	.080	72.6
278	.007	.006	.029	.003	.012	.011	.009	.008	.010	.005	.012	17.3
285	.004	.012	.029	.023	.006	.013	.011	.014	.013	.003	.003	15.0
290	.004	.008	.029	.031	.012	.015	.011	.021	.017	.021	.006	62.8
298	.002	.007	.029	.026	.006	.007	.007	.010	.011	.005	.037	45.0
310	.005	.013	.029	.049	.012	.029	.013	.019	.029	.021	.025	119.2
331	.051	.024	.029	.028	.017	.024	.012	.026	.019	.058	.037	86.1
332	.005	.007	.029	.018	.003	.012	.005	.013	.012	.003	.012	33.8
342	.027	.148	.059	.056	.093	.260	.192	.202	.189	.026	.160	532.5
344	.002	.013	.059	.005	.006	.025	.015	.015	.018	.005	.012	58.3
345	.006	.011	.029	.031	.017	.009	.012	.012	.010	.003	.006	27.2
401	.242	.292	.059	.138	.297	.211	.185	.252	.215	.238	.196	270.7
402	.042	.054	.029	.013	.070	.071	.050	.055	.042	.026	.012	138.6
410	.014	.009	.029	.005	.012	.014	.033	.015	.014	.005	.003	105.0
412	.009	.010	.029	.005	.012	.017	.028	.021	.017	.016	.006	58.3
413	.007	.010	.029	.003	.023	.020	.021	.019	.017	.011	.012	34.4
414	.059	.088	.059	.077	.081	.174	.189	.132	.126	.079	.049	378.9
424	.014	.011	.029	.008	.023	.016	.045	.017	.013	.016	.003	224.2
426	.012	.018	.059	.015	.012	.023	.023	.026	.020	.016	.018	32.6
427	.129	.144	.059	.082	.198	.136	.483	.165	.143	.042	.098	2851.6
428	.040	.052	.059	.056	.052	.085	.124	.079	.087	.016	.061	189.7
429	.019	.032	.029	.026	.029	.048	.060	.045	.050	.003	.025	76.3
433	.014	.011	.029	.015	.012	.029	.061	.050	.026	.074	.031	391.0
434	.021	.020	.029	.023	.012	.018	.015	.003	.021	.026	.018	231.9
435	.017	.005	.029	.018	.006	.014	.026	.023	.019	.053	.098	156.0
436	.017	.033	.029	.031	.064	.022	.021	.019	.003	.037	.123	630.0
437	.047	.047	.029	.028	.047	.071	.032	.054	.042	.048	.067	145.2
438	.007	.024	.118	.026	.003	.025	.021	.031	.025	.053	.190	218.8
440	.011	.032	.029	.008	.023	.054	.032	.036	.043	.016	.018	103.9
443	.002	.007	.029	.008	.006	.013	.011	.011	.011	.005	.012	19.6
486	.032	.031	.029	.018	.035	.029	.019	.027	.034	.005	.006	46.8
496	.037	.043	.059	.038	.023	.043	.039	.049	.044	.063	.018	17.7
507	.027	.025	.029	.008	.058	.016	.017	.019	.017	.005	.006	44.3
599	.075	.090	.029	.067	.076	.092	.087	.102	.100	.063	.049	34.3
715	.010	.020	.029	.013	.017	.034	.021	.024	.029	.042	.025	56.2
780	.090	.086	.059	.072	.087	.051	.058	.069	.060	.085	.074	91.2
784	.025	.045	.059	.028	.017	.088	.090	.085	.067	.069	.049	190.0
787	.002	.009	.029	.001	.003	.014	.008	.012	.013	.016	.006	25.9
788	.009	.010	.029	.018	.023	.011	.007	.012	.011	.005	.031	18.8

NOTE: Each entry is the proportion of patients with the secondary diagnosis given by the row code from among those patients with principal diagnosis given by the column code. A large value for the chi square statistic is 20 under the null hypothesis of no differences.

diagnoses is not strong enough to explain all of the variation in posthospital use among patients with different secondary diagnoses.

In analogy with the standard regression analysis, the additional predictive capability of all of the secondary diagnoses, $S_1,...,S_{n_s}$, may be due in large part to the presence of a small subset of the secondary diagnoses. This would be of interest if it

were true because one of the primary purposes of this Note is to examine the feasibility of explaining the variability in posthospital use by a relatively small number of categories such as those used in the hospital DRG classification system. There is no natural ordering of the secondary diagnoses for performing a sequential analysis for the inclusion of some, but not all secondary diagnoses. An exploratory analysis was performed instead to examine all of the distinct pairs of secondary diagnoses to determine if some pairs are heavily dependent on each other. This dependence could come in the form of pairs of diagnoses that frequently occur together, or pairs that never occur together.

A two-way table was formed for each pair of secondary diagnoses, and the Yule Q measure of concordance (Bishop et al., 1975) was calculated. This statistic is close to one when two secondary diagnoses occur together with high frequency, and near negative one when they seldom occur together. These statistics show that the secondary diagnoses are much more highly correlated than could be explained by sampling fluctuations under a null model of independence. The nature and magnitude of the dependence is similar for each pair. The dependence occurs because when any one of the secondary diagnoses is present, the odds that any one of the other secondary diagnoses will also be present are substantially increased. Despite this strong dependence, no specific pairs of secondary diagnoses were linked so that the presence (or absence) of one determined the presence (or absence) of the other. As a result, no small subset of the secondary diagnoses can be used to predict the values of all secondary diagnoses.

The dependence between secondary diagnoses may have two sources. The first is that the presence of a secondary diagnosis may indicate that the patient is severely ill and is more likely to have other problems as well. The second explanation is that the reporting of secondary diagnoses in the records we are using is voluntary; no direct reimbursement decisions are based on secondary diagnoses. Anecdotal evidence suggests that reporting practices may vary substantially among doctors and hospitals. The presence of any secondary diagnosis may then simply serve as an indicator that the patient was treated at an institution that was more diligent in recording secondary diagnoses, and may not be indicative of any special medical condition. This latter possibility is important because it also provides an alternative interpretation of the predictive capability of secondary diagnoses. If some institutions record secondary diagnoses much more frequently than others, and these institutions also send patients into posthospital facilities more often, then the fact that patients with reported secondary diagnoses use posthospital care more often may be a consequence of reporting patterns rather than different patient characteristics. Although this alternative interpretation of our

data does cast doubt on the replicability of the estimated models using secondary diagnoses obtained from some other data source, it does not account for the substantial differences in posthospital care use for patients with different secondary diagnoses. Although reporting inconsistencies in our data complicate the interpretation of the predictive value of secondary diagnoses measured from our data, differences in posthospital use rates among patients with different secondary diagnoses still appear to be associated with different physical characteristics of individual patients.

Forming the two-way tables for all of the different pairs of secondary diagnoses requires expensive computer tabulation. Although the principal diagnoses are very different in each DRG, the more frequently occurring secondary diagnoses tend to be common to all of the DRG categories. Because the same secondary diagnoses are prominent for each DRG, we conjecture that the relationship between secondary diagnoses does not change much for different DRG categories. In light of this opinion, and the expense involved in forming the two-way tables, this analysis was not repeated for other DRG categories.

III. MODELS FOR POSTHOSPITAL UTILIZATION

Models of the probability of posthospital use as a function of a patient's principal and secondary diagnoses are described in this section. Formal statistical evidence is presented supporting the claim that secondary diagnoses contain additional predictive information beyond that contained in the principal diagnoses.

In its most basic formulation, the question being addressed is whether patients with the same principal diagnosis but different secondary diagnoses use posthospital care at different rates. The purpose of the modeling exercise documented in this section is to explore the enormous number of different combinations of principal and secondary diagnoses to determine if there are many examples where patients with the same principal diagnosis but different secondary diagnoses receive substantially different amounts of posthospital care. In addition to providing an efficient way to examine many subsets of patients, this formalized modeling process provides a method for quantifying the size of differences in posthospital utilization rates allowing comparison of the relative importance of the principal and secondary diagnoses. With so many combinations of principal and secondary diagnoses, the sample sizes for each subgroup are fairly small, so differences in the observed rates are to be expected, because of sampling fluctuations, even in the absence of any mechanism linking diagnoses and posthospital use. The modeling done in this section also provides standard statistical tools for assessing the extent to which the differences in observed use rates are due to our selection of these larger sampling fluctuations.

A GENERAL DESCRIPTION OF THE MODEL

The notation introduced here is common to all models of the different types of posthospital care, and for the other DRG categories summarized in Appendix A. Let Y be a 0/1 variable that represents the use of posthospital care. The three types of posthospital care, HHA, SNF, and REHAB, are treated separately, with Y representing use of each type of care. The representation of the principal and secondary diagnoses introduced in Sec. II is reused throughout this section.

The model that was fit assumes an additive contribution from each principal and secondary diagnosis to the log odds (logit) of the probability that the patient uses posthospital care. The model is given by

$$\text{Logit}\left(P(Y=1 \mid D_1,\ldots D_{n_d}, S_1,\ldots S_{n_s})\right) = \alpha_0 + \alpha_1 D_1 + \cdots + \alpha_{n_d} D_{n_d} + \beta_1 S_1 + \cdots + \beta_{n_s} S_{n_s} + \gamma I \quad . \qquad (1)$$

Several features of this model are worth noting. If the S_i are excluded, the model just estimates a patient's chance of using a posthospital service as the proportion of all patients with the same principal diagnosis who receive the posthospital care,

$$\text{Logit}\left(P(Y=1 \mid D_1,\ldots D_{n_d})\right) = \alpha_0 + \alpha_1 D_1 + \cdots + \alpha_{n_d} D_{n_d} \quad . \qquad (2)$$

These proportions, given as percentages, are in Table 1.1. This case involves few modeling assumptions and yields very similar results regardless of the exact form of the model. The inclusion of the indicators for secondary diagnoses, however, adds substantial complexity to the model and requires strong assumptions which are at best only approximately correct. The most important and also most dubious modeling assumption is that the presence of each secondary diagnosis changes the log odds of posthospital care use by the same amount regardless of which principal diagnosis is present, and regardless of whether there are other secondary diagnoses. This is a crude approximation, since the change in odds given the presence of a secondary diagnosis might be different depending on the patient's principal diagnosis. This situation is referred to as an *interaction* between the secondary and principal diagnoses. When interactions are present, the model can merely indicate the direction of the changes in odds for a specific secondary diagnosis associated with each of the principal diagnoses. Furthermore, the model can be accurate only insofar as most of these changes are of about the same size and direction. The model also assumes that the change in odds for secondary diagnosis S_i is the same when it is the only secondary diagnosis as when there are several secondary diagnoses present. This assumption becomes more dubious as the number of secondary diagnoses recorded for a patient increases. In particular, in the data used to estimate the model, at most four secondary diagnoses can be present for a patient, and it would be foolish to extrapolate the model to make predictions when there are more than four secondary diagnoses present. The likelihood ratio statistic comparing models (1) and (2) provides a statistic for formally testing the hypothesis that the secondary diagnoses are not predictive of posthospital use after accounting for principal diagnosis.

Models that allow for all or many of the interaction terms are possible in principle, but they are not practical because there are thousands of possible interactions

for the large number of diagnoses being considered. Not only would this cause computational difficulty, but even with the large data set available, most interaction estimates would be based on few or no data. As a crude approximation, a variable was added to model (1) to indicate whether there was more than one secondary diagnosis present for a patient. This is a type of *average interaction* among the secondary diagnoses. It was included to detect whether there may be a consistent negative interaction term among secondary diagnoses corresponding to the situation where the presence of any secondary diagnosis raises the odds of posthospital use, but the presence of additional reported secondary diagnoses is of little relevance. Let I represent this common interaction term. The variable I is 1 if more than one secondary diagnosis is present, and zero otherwise. Model (1) is then supplemented by I to give

$$\begin{aligned}\text{Logit}\left(P\left(Y = 1 \mid D_1, \ldots D_{n_d}, S_1, \ldots S_{n_s}\right)\right) &= \alpha_0 + \alpha_1 D_1 + \cdots + \alpha_{n_d} D_{n_d} \\ &+ \beta_1 S_1 + \cdots + \beta_{n_s} S_{n_s} + \gamma I \quad . \end{aligned} \tag{3}$$

This average interaction term had small estimated values for most of the outcomes and DRG categories that were modeled. In many cases, the interaction has a positive value indicating that individuals with many secondary diagnoses may be even more likely to use posthospital care.

ESTIMATING THE LOGISTIC MODELS

The large number of parameters that must be estimated, and the large sample sizes available (and necessary to provide accurate parameter estimates with so many parameters), presented computational difficulties that were resolved by fitting the models in a two-step process. For the first step, the complete model in (1) was estimated using the sample linear discriminant estimate of the α's and β's. This estimator was obtained by applying a very simple modification to the standard least squares regression estimator of model (1), where the *response* variable is the 0/1 variable indicating whether each patient used a particular type of posthospital care. Haggstrom (1983) discusses the use of this easy-to-compute estimator as a very good approximation to the iterative logistic regression maximum likelihood estimator. The estimated parameters of the secondary diagnoses were then compared with their estimated standard errors by forming t-statistics. A subset of 13 secondary diagnoses whose t-statistics showed the most evidence of changing the odds of posthospital care use were retained for the next step in the estimation process. The estimation was then

completed using the logistic regression maximum likelihood estimates of the parameters for the subset model.

Reducing the number of secondary diagnoses included in the model has two benefits. First, it reduces the computational difficulty and allows logistic regression estimators to be calculated. Second, and more important, the smaller subset is easier to explore and comprehend. All of the principal diagnoses are included in the resulting reduced models. The number of secondary diagnoses retained in the reduced models depended on the sample size (for computational reasons), and varies by a small amount for different diagnoses. In addition to the secondary diagnoses with large parameter estimates, a new variable was formed to indicate when a patient has a secondary diagnosis (including those in the original other category) that is not among those in the reduced list.

A major concern arising from an estimation process like this is that the magnitude of the parameter estimates in the final models, and the corresponding inferential quantities such as the t-statistics and likelihood ratio statistics, will be biased in the direction of exaggerating the predictive capability of the secondary diagnoses. To get some indication of the size of this bias, the 51,165 patients in DRG 14 were divided into two random, mutually exclusive subsets of roughly equal size. Estimates based on the first subset are presented in the remainder of this section. The same procedures were applied to the second half of the data with the results presented in Appendix B for comparison. The number of patients in DRG 14 is roughly double the number of patients in the other DRG categories considered in Sec. IV, so the differences between the two subsets should indicate the stability of the estimation process for the other DRG categories. For most posthospital outcome variables and DRG categories, the magnitudes of the parameter estimates and inferential statistics are much larger than the effect resulting from selecting only a subset of the secondary diagnoses.

The estimates of the parameters for models (1) and (3) for the HHA outcome variable are given in Table 3.1. The estimates for model (1), which include all of the secondary diagnoses, were fit using the sample linear discriminant and are labeled "Full." Estimates obtained using logistic regression for a model like that in (3), but which contains only the 13 secondary diagnoses with the largest t-statistics, are labeled "Reduced." In general, there is good agreement between the parameter estimates common to both models. This is because the logistic regression estimator and the

Table 3.1

ESTIMATED MODELS FOR DRG 14: HHA

Code	Full Est.	Full t-stat.	Reduced Est.	Reduced t-stat.
		Principal Diagnoses		
430	-0.42	-3.43	-0.62	-3.38
431	-0.09	-0.44	-0.12	-1.77
4320	-1.28	-1.67	-0.88	-0.82
4321	-0.67	-3.19	-0.99	-3.84
4329	-0.35	-0.31	-0.13	-0.38
4340	0.10	3.01	0.09	2.10
4341	0.13	1.92	0.06	0.98
4349	0.13	4.36	0.08	2.32
4373	-0.83	-3.14	-1.20	-3.25
7843	-0.42	-1.40	-0.68	-2.10
		Secondary Diagnoses		
V10	0.04	0.31	—	—
V45	-0.06	-0.54	—	—
41	-0.01	-0.07	—	—
244	0.17	1.26	—	—
250	0.21	5.33	0.19	4.69
276	-0.05	-0.82	—	—
278	-0.18	-1.28	—	—
285	0.18	1.45	—	—
290	-0.20	-1.84	-0.21	-1.80
298	-0.02	-0.15	—	—
310	-0.09	-1.01	—	—
331	0.07	0.75	—	—
332	0.05	0.41	—	—
342	0.23	6.49	0.23	6.06
344	0.32	2.92	0.30	2.90
345	0.01	0.08	—	—
401	0.14	4.21	0.13	3.41
402	0.18	2.68	0.16	2.37
410	0.05	0.39	—	—
412	0.06	0.57	—	—
413	0.08	0.74	—	—
414	-0.05	-1.07	—	—
424	-0.08	-0.71	—	—
426	0.16	1.67	—	—
427	-0.05	-1.23	—	—
428	0.10	1.76	—	—
429	0.04	0.60	—	—
433	-0.14	-1.92	-0.15	-1.97
434	0.30	2.58	0.28	2.43
435	-0.14	-1.36	—	—
436	0.24	1.97	—	—
437	0.02	0.23	—	—
438	0.12	1.33	—	—
440	-0.00	-0.04	—	—
443	0.06	0.41	—	—
486	-0.34	-3.47	-0.37	-3.40
496	-0.02	-0.27	—	—
507	-0.30	-2.37	-0.33	-2.37
599	0.12	2.10	0.11	2.15
715	-0.00	-0.05	—	—
780	-0.23	-3.72	-0.26	-3.77
784	0.04	0.67	—	—
787	-0.11	-0.86	—	—
788	0.39	2.91	0.35	2.82
Other	-0.02	-0.64	-0.04	-0.90
Interact	—	—	0.02	0.46

sample linear discriminant are similar estimators for most types of data, and because the correlation between diagnoses is low as indicated in Sec. II.

The estimates for the principal diagnoses given at the beginning of the table require some additional explanation. Because there were only about 25 patients in DRG 014 with a principal diagnosis not listed in Table 1.1, these patients were excluded from the analysis. To provide a more appropriate reference category for the principal diagnoses estimates, the principal diagnosis with the largest number of patients, 436 CVA, was used as the reference in model (1). It is indicated by setting all of the principal diagnosis indicator variables to zero. The parameter estimates for principal diagnoses measure differences with principal 436, but differences between other primaries can be obtained directly from these estimates by subtracting the parameter estimates.

Judging by the size of the differences in the parameter estimates, the use of HHA care tends to vary more with different principal diagnosis than with different secondary diagnosis. This is consistent with the unadjusted percentages given in Tables 1.1 and 1.2. Judging from the t-statistics, there is still evidence that HHA use varies among patients with different secondary diagnoses. As indicated by the results in Appendix B, the larger t-statistics will shrink somewhat as a result of selection effects when the estimation process is repeated with an independent replication of the sample, but this shrinkage does not account for most of the larger values. The change in odds predicted for patients with different secondary diagnoses can be substantial. For example, a patient with diagnosis 788, symptoms involving the urinary system, is predicted to be about one and a half times as likely to use HHA care as a patient with no secondary diagnoses, and twice as likely to use HHA care as a patient with diagnosis 486, pneumonia. Although results such as these must be viewed cautiously because of the crude nature of the statistical models and the selection effects that occur when a small number of significant parameters are selected from a large group of parameters, the analyses presented here and in Appendix B support the claim that secondary diagnoses provide meaningful information for predicting HHA use beyond that contained in the principal diagnosis.

The logistic regression procedure provides a formal statistical test of this hypothesis. The reduced model in Table 3.1, which includes secondary diagnoses, can be compared with the estimate of model (2). A test that the parameters associated with the secondary diagnoses are all zero is obtained from the likelihood ratio statistic for comparing these nested models. Under the null hypothesis, twice the difference in the

log likelihood ratios for these models should be approximately a chi square random variable with degrees of freedom equal to the difference in the number of parameters in the models. Twice the log likelihood ratio for the HHA models is 175.98, and the difference in the number of parameters is 15, the number of secondary diagnoses plus the "other" and "interaction" parameters. The test statistic is highly significant. The degrees of freedom here are too small, however, since the 13 secondary diagnoses were selected from a list of 44 secondary diagnoses. A conservative way to correct for this selection effect is to assume that the number of additional parameters is actually that of the larger model with all secondary diagnoses, but even with 50 degrees of freedom, the test statistic values are still far from plausible under the null hypothesis.

Note that the role of the additive logistic models in Eqs. (1) and (3) is only to specify alternatives to the null hypothesis that only principal diagnoses provide information about posthospital use. Although the ability of the test to detect deviations from this hypothesis depends on the choice of models in (1) and (3), the validity of the reported P-values does not depend on these models being correctly specified. The much more important question of how much additional information is contained in the secondary diagnoses, however, does depend on the models specified in (1) and (3).

The corresponding results for SNF use are given in Table 3.2. The parameter estimates tend to be larger and have larger t-statistics than for HHA use indicating a stronger dependence on the diagnoses. With the exception of the two rare principal diagnoses, the parameter estimates for the principal diagnoses are of approximately the same magnitude as those for the secondary diagnoses, indicating that differences in SNF use for patients with different secondary diagnoses is about the same as that for patients with different principal diagnoses. It is interesting to note that the interaction term has a positive estimate that is borderline significant. The presence of a negative interaction term along with several substantially negative parameter estimates for some of the secondary diagnoses makes this term particularly difficult to interpret. When it is removed from the model, the remaining parameter estimates change very little because of the low correlation of each secondary diagnosis with the interaction term. The likelihood ratio statistic, computed in the same way as described for the HHA variable, is 729.42. This is highly significant regardless of the choice of plausible degrees of freedom.

Table 3.2

ESTIMATED MODELS FOR DRG 14: SNF

	Full		Reduced	
Code	Est.	t-stat.	Est.	t-stat.
Primary Diagnoses				
430	-0.09	-0.06	-0.21	-0.94
431	0.51	5.62	0.48	6.74
4320	0.12	0.08	0.16	0.15
4321	-0.36	-1.71	-0.71	-2.16
4329	-0.03	-0.34	-0.12	-0.27
4340	0.07	3.18	0.14	2.48
4341	-0.27	-1.76	-0.18	-2.22
4349	0.13	3.31	0.17	3.98
4373	-0.44	-1.26	-0.97	-1.90
7843	0.01	0.02	-0.24	-0.59
Secondary Diagnoses				
V10	0.23	1.57	—	—
V45	0.09	0.65	—	—
41	0.73	6.64	0.51	5.54
244	0.51	2.99	—	—
250	0.13	2.59	—	—
276	0.45	5.87	0.39	5.68
278	-0.1	-1.07	—	—
285	0.17	1.07	—	—
290	0.37	2.72	—	—
298	0.12	0.67	—	—
310	0.27	2.34	—	—
331	-0.0	-0.05	—	—
332	0.32	1.88	—	—
342	0.63	13.83	0.50	11.31
344	0.44	3.19	—	—
345	-0.0	-0.50	—	—
401	-0.0	-0.72	—	—
402	0.15	1.85	—	—
410	0.11	0.65	—	—
412	-0.3	-2.48	-0.47	-2.89
413	-0.2	-1.74	—	—
414	0.05	0.96	—	—
424	0.09	0.66	—	—
426	-0.1	-1.11	—	—
427	0.44	8.20	0.37	7.54
428	0.38	5.40	0.3	5.00
429	0.12	1.44	—	—
433	-0.3	-3.78	-0.55	-4.62
434	0.06	0.43	—	—
435	-0.5	-3.90	-0.82	-4.44
436	0.48	3.04	—	—
437	-0.00	-0.04	—	—
438	-0.0	-0.38	—	—
440	0.04	0.38	—	—
443	-0.1	-0.58	—	—
486	0.84	6.79	0.64	6.23
496	0.11	1.27	—	—
507	1.41	8.84	0.96	7.92
599	0.48	6.92	0.39	6.17
715	0.11	1.03	—	—
780	0.22	2.79	—	—
784	0.36	5.30	0.28	4.66
787	1.15	7.11	0.82	6.56
788	0.51	3.00	—	—
Other	0.10	2.73	0.09	1.60
Interact	—	—	0.11	2.20

The results for the REHAB variable are given in Table 3.3. They are very similar to those for SNF use and no additional comments are included here. The likelihood ratio statistic for the null hypothesis of additional information in the secondary diagnoses for REHAB is highly significant with a value of 635.54.

DIFFERENTIATING BETWEEN PATIENTS USING ESTIMATED MODELS

To quantify the ability of the estimated models to differentiate between patients with high and low posthospital utilization, the models estimated in the previous section were used to compute an estimated probability of (each type of) posthospital use for the half of the data excluded when the models were selected and estimated. For each patient in the excluded data, the estimated probability of posthospital use was calculated, and then these patients were divided into ten equal-sized groups with increasing estimated probabilities of posthospital use. Within each of these deciles of estimated probabilities, the proportion of patients who actually used posthospital care was also computed. This process was performed for each type of posthospital use separately. The results, shown in Table 3.4, confirm that the models divide the patients into groups of substantial size that have markedly different levels of posthospital use. The average estimated probabilities in the extreme categories tend to be even more extreme than the actual observed use rates. Some degradation in the model predictions of this type was anticipated; this corresponds to the shrinkage of large t-statistics discussed in the previous section.

There are sizeable differences in the patients grouped by their predicted values, considerably more than can be obtained using only the principal diagnosis in Table 1.1 to distinguish between patients. The improvements over Table 1.1 are even larger than first appearances suggest because the principal diagnoses with the more extreme values are those that occur very rarely, so their estimated posthospital use is much more variable, and in any case, these diagnoses do not help much in distinguishing between patients because they occur so infrequently. A more direct comparison of the results in Table 1.1 with the categories formed by the predicted values in Table 3.4 was obtained by computing the posthospital use rates for each principal diagnosis using only the training set that was also used to estimate the models with secondary diagnoses, and then these rates were used to predict the rates for the data excluded from the fitting process. Since there are only a few principal diagnoses, and the patients are highly concentrated in only five of the principal diagnoses, deciles were formed by pooling

Table 3.3

ESTIMATED MODELS FOR DRG 14: REHAB

	Full		Reduced	
Code	Est.	t-stat.	Est.	t-stat.
Primary Diagnoses				
430	-0.26	-1.00	-0.67	-1.60
431	0.54	3.17	0.34	3.24
4320	-0.72	-0.51	-0.53	-0.41
4321	-0.15	-0.38	-0.87	-1.48
4329	0.09	0.42	-0.30	-0.40
4340	0.13	1.90	0.13	1.70
4341	0.38	3.50	0.45	4.63
4349	0.46	7.56	0.38	6.52
4373	-0.62	-1.65	-0.70	-0.98
7843	-0.23	-0.06	-0.16	-0.27
Secondary Diagnoses				
V10	0.02	0.10	—	—
V45	0.29	1.51	—	—
41	-0.04	-0.29	—	—
244	0.50	2.13	—	—
250	0.18	2.64	0.11	1.64
276	-0.30	-2.88	-0.46	-3.54
278	-0.31	-1.25	—	—
285	-0.45	-2.06	—	—
290	-0.59	-3.20	-1.36	-3.78
298	-0.42	-1.79	—	—
310	-0.49	-3.16	-0.98	-3.82
331	-0.30	-1.81	—	—
332	-0.05	-0.24	—	—
342	1.29	20.70	0.94	16.33
344	-0.05	-0.25	—	—
345	-0.39	-1.65	—	—
401	0.39	6.63	0.30	5.05
402	0.27	2.33	0.20	1.86
410	0.96	4.09	0.67	3.75
412	0.41	2.22	—	—
413	0.23	1.28	—	—
414	-0.17	-2.21	-0.16	-2.09
424	0.20	1.03	—	—
426	0.09	0.50	—	—
427	0.14	1.84	—	—
428	-0.11	-1.16	—	—
429	-0.29	-2.44	-0.37	-2.76
433	-0.22	1.73	—	—
434	0.15	0.77	—	—
435	-0.44	-2.53	-0.70	-2.81
436	0.0	0.02	—	—
437	-0.15	-1.26	—	—
438	-0.08	-0.50	—	—
440	-0.17	-1.34	—	—
443	0.22	0.97	—	—
486	-0.32	-1.90	—	—
496	-0.01	-0.06	—	—
507	0.07	0.31	—	—
599	0.05	0.49	—	—
715	0.03	0.22	—	—
780	-0.35	-3.33	-0.53	-3.91
784	0.53	5.68	0.33	4.44
787	-0.02	-0.11	—	—
788	0.26	1.12	—	—
Other	0.12	2.33	-0.02	-0.25
Interact	—	—	0.18	2.24

Table 3.4

PREDICTED AND OBSERVED POSTHOSPITAL UTILIZATION RATES

HHA		SNF		REHAB	
Observed	Predicted	Observed	Predicted	Observed	Predicted
0.210	0.189	0.077	0.079	0.022	0.018
0.239	0.226	0.105	0.102	0.034	0.035
0.248	0.242	0.105	0.114	0.043	0.038
0.264	0.253	0.112	0.124	0.054	0.047
0.273	0.262	0.141	0.128	0.053	0.053
0.271	0.275	0.155	0.140	0.061	0.060
0.278	0.285	0.148	0.150	0.076	0.071
0.295	0.303	0.185	0.167	0.092	0.090
0.300	0.317	0.210	0.187	0.113	0.121
0.325	0.359	0.285	0.258	0.176	0.187

NOTE: The predicted values are based on the models estimated in Sec. III. The observed and predicted values are for the half of the data excluded from the model fitting.

principal diagnoses with similar posthospital use and randomly selecting patients from large diagnoses that must be split to obtain deciles. The upper and lower extreme deciles based on the principal diagnoses for HHA are 0.220 and 0.292, for SNF the deciles are 0.124 and 0.192, and for REHAB, the deciles are 0.050 and 0.088. Comparing these extreme deciles to those in Table 3.4 shows that patients with much higher rates of posthospital care can be selected from within the DRGs, and substantial further differentiation of patients is possible if secondary diagnoses are considered.

SUMMARY OF RESULTS FOR OTHER DRG CATEGORIES

This section summarizes the results from the other DRG categories, with emphasis on comparisons of posthospital use among secondaries occurring in each of the five DRG categories studied. More complete results for the remaining DRG categories are given in Appendix A. The main purpose of this analysis is to identify groups of secondary diagnoses that have consistent trends across different DRG categories and principal diagnoses. The existence of such secondary diagnoses lends some credibility to the modeling process and the use of the secondary diagnoses recorded in the Medicare administrative records. More important, they may be helpful in identifying secondary diagnoses, or groups of secondary diagnoses, for further study using data with more extensive medical information.

As part of this analysis, we consulted a clinician experienced in the treatment of Medicare patients and the assignment of these patients to posthospital care. Each

secondary diagnosis was assigned one of three ratings for each type of posthospital care: more likely to use care, less likely to use care, or not related to care decisions. These subjective ratings, based on medical experience, were then compared with the results estimated from the Medicare data, and although there were a few disagreements (which will be discussed below), the estimated results were generally consistent with clinical experience. As was expected, most secondary diagnoses do not appear to be very strongly related to SNF use, but for those that are, the tendency is to increase the chance of SNF use above that of patients with no recorded secondary diagnoses. The same was expected of HHA use, and although this was generally observed in the estimated results, there were more exceptions to this for HHA. An explanation of why the presence of some secondary diagnoses is associated with less HHA use will be given below. The results for REHAB use are more mixed than expected, since some conditions make it unlikely that the the patient will show the demonstrated improvement necessary for the REHAB to receive reimbursement. A much smaller number of secondary diagnoses were judged to be significantly related to REHAB use than the other types of posthospital care. This is almost certainly a consequence of the fact that the rates of REHAB use are so small that it is very difficult to detect differences even if they represent a doubling or tripling of the odds of REHAB use. No systematic attempt was made to quantify the extent of agreement between the clinician's judgment and the results of the estimation procedure. The goal of this analysis was to obtain a general impression of the clinical reasonableness of the estimated models, and to identify unusual predictions from the estimated models that warrant further examination. A detailed medical interpretation of each parameter in the estimated models is beyond the scope of this analysis.

The HHA parameter estimates and their t-statistics for the secondary diagnoses common to all five of the DRG categories are given in Table 3.5. A number of these secondaries showed consistent trends toward increased HHA use including diabetes (250), anemias (285), hypertension and related heart disease (401, 402), and urinary tract problems (599). Patients with psychological diseases, especially those related to identifiable brain degeneration, tended to use less HHA care. These secondaries are not represented in Table 3.5 (except for diagnosis 310) because they are not common to all secondary diagnoses. We speculate that the reduced use of HHA care among these patients results because these patients are so ill that they are seldom discharged to home care. If this is the case, the models would be correct in identifying these patients, but our interpretation of the negative estimates would be much different. To check if this is

happening, we estimated an identical model for DRG 209, which had several large negative estimates, using only patients who were discharged to home care according to the information contained in the MEDPAR data. Although the estimated parameters for these secondaries did decrease in magnitude, they remain negative with significant t-statistics, warranting further study. One possible explanation for this observation is that the MEDPAR information indicating discharge to home care may fail to reflect the actual status of these patients because many of these patients may in fact be institutionalized, but not under Medicare funding, so their institutional status is not reported in the MEDPAR files.

Table 3.5

SECONDARY DIAGNOSES COMMON TO ALL DRG CATEGORIES: HHA USE

	DRG 14		DRG 88		DRG 127		DRG 209		DRG 210	
Code	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
V10	0.036	0.308	0.019	0.194	-0.046	-0.376	0.044	0.517	0.163	1.702
V45	-0.060	-0.540	-0.212	-1.518	-0.047	-0.603	-0.028	-0.238	0.265	1.943
41	-0.006	-0.067	0.019	0.230	-0.017	-0.147	-0.119	-1.508	0.021	0.27
244	0.170	1.257	0.220	1.531	0.186	1.469	-0.134	-1.309	0.267	1.943
250	0.210	5.326	0.194	3.282	0.149	3.785	0.161	3.123	0.099	1.853
276	-0.050	-0.823	0.279	4.743	0.179	2.830	-0.007	-0.096	-0.125	-1.864
285	0.183	1.449	0.518	3.975	0.117	1.454	0.177	3.337	0.028	0.529
310	-0.091	-1.005	0.080	0.579	-0.050	-0.399	-0.746	-8.317	-0.894	-13.2
401	0.142	4.211	-0.059	-1.146	-0.042	-0.855	0.106	3.160	0.243	5.469
402	0.178	2.681	0.018	0.153	-0.044	-0.503	0.191	2.177	0.233	2.262
412	0.061	0.570	-0.144	-1.202	-0.099	-1.317	-0.021	-0.180	0.113	0.787
413	0.078	0.743	0.064	0.817	-0.055	-0.851	0.430	4.316	0.110	0.894
414	-0.047	-1.074	0.010	0.230	-0.057	-1.707	0.055	1.111	-0.069	-1.381
426	0.164	1.670	-0.064	-0.540	-0.122	-1.740	-0.114	-1.307	0.166	1.863
427	-0.052	-1.226	0.269	5.440	0.030	0.864	0.093	1.719	-0.082	-1.516
428	0.099	1.759	0.462	11.221	-0.055	-0.362	0.014	0.193	-0.241	-3.933
429	0.041	0.599	-0.002	-0.022	0.011	0.196	0.040	0.566	0.197	-2.738
440	-0.003	-0.044	0.233	2.011	0.016	0.179	-0.172	-1.789	-0.163	-1.885
496	-0.019	-0.272	0.217	2.160	0.361	6.237	0.056	0.874	0.854	1.407
599	0.115	2.101	0.490	5.555	0.103	1.234	0.208	3.483	-0.093	-1.702
715	-0.004	-0.050	0.093	1.007	-0.027	-0.214	0.097	1.736	0.646	0.879
780	-0.228	-3.723	0.015	0.149	0.428	3.058	0.099	0.954	-0.046	-0.485

Table 3.6 shows the parameter estimates and t-statistics for SNF use just as those in Table 3.5. The secondary diagnoses with strong positive trends across the five DRG categories are fluid disorders (276), anemias (285), organic mental disorders (310), heart diseases (414, 427, 428), and urinary tract problems (599). Patients with

Table 3.6

SECONDARY DIAGNOSES COMMON TO ALL DRG CATEGORIES: SNF USE

	DRG 14		DRG 88		DRG 127		DRG 209		DRG 210	
Code	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
V10	0.228	1.568	-0.675	-2.150	-0.294	-0.897	-0.119	-1.017	0.080	0.820
V45	0.091	0.651	-0.223	-0.505	0.007	0.032	-0.027	-0.168	0.024	0.170
41	0.729	6.642	0.229	0.896	0.556	1.800	-0.059	-0.543	-0.027	-0.339
244	0.509	2.989	0.227	0.499	0.170	0.498	0.048	0.345	0.083	0.588
250	0.129	2.586	-0.060	-0.321	-0.264	-2.500	0.157	2.226	-0.019	-0.338
276	0.450	5.874	0.780	4.189	1.071	6.303	0.452	4.651	0.169	2.446
285	0.170	1.068	0.867	2.098	0.069	0.320	0.147	2.023	0.115	2.096
310	0.267	2.344	1.890	4.339	1.951	5.827	0.942	7.681	0.135	1.939
401	-0.031	-0.718	-0.258	-1.570	-0.230	-1.736	-0.109	-2.362	-0.030	-0.650
402	0.155	1.846	-0.307	-0.842	-0.472	-2.004	0.256	2.135	0.317	2.999
412	-0.335	-2.478	-0.578	-1.526	-0.045	-0.225	-0.111	-0.702	0.009	0.063
413	-0.232	-1.738	-0.493	-1.977	-0.362	-2.089	-0.271	-1.983	0.226	1.781
414	0.053	0.959	0.077	0.553	-0.086	-0.963	0.393	5.851	0.147	2.880
426	-0.137	-1.106	-0.787	-2.108	-0.205	-1.090	0.014	0.115	-0.103	-1.124
427	0.441	8.201	0.531	3.383	0.325	3.462	0.392	5.305	0.309	5.561
428	0.382	5.396	0.467	3.577	0.448	1.092	0.568	5.765	0.146	2.316
429	0.124	1.441	-0.492	-1.961	-0.154	-1.046	0.092	0.946	0.141	1.914
440	0.036	0.384	-0.020	-0.054	0.314	1.319	0.198	1.507	0.034	0.379
496	0.110	1.269	0.310	0.972	0.668	4.295	0.191	2.174	-0.047	-0.760
599	0.480	6.916	1.233	4.415	0.144	0.641	0.290	3.548	0.134	2.392
715	0.108	1.028	0.025	0.086	0.302	0.879	0.000	-0.006	0.063	0.833
780	0.216	2.793	0.984	2.987	-0.540	-1.431	-0.072	-0.511	0.052	0.541

psychological diseases related to identifiable brain degeneration were also consistently high users of SNF care.

As anticipated, the presence of brain diseases, 290, 310, and 331, decreased the use of REHAB care. The presence of Parkinson's disease in DRGs 209 and 210, however, is an unexplained exception to this trend. The estimates for REHAB care for the common secondary diagnoses are given in Table 3.7.

Table 3.7

SECONDARY DIAGNOSES COMMON TO ALL DRG CATEGORIES: REHAB USE

	DRG 014		DRG 209		DRG 210	
Secondary Code	Parameter Estimate	t-stat.	Parameter Estimate	t-stat.	Parameter Estimate	t-stat.
V10	0.020	0.103	0.082	0.301	0.155	0.565
V45	0.290	1.511	0.316	0.849	-0.011	-0.029
41	-0.044	-0.291	-0.037	-0.148	-0.247	-1.104
244	0.497	2.134	-0.234	-0.718	0.220	0.556
250	0.180	2.642	0.166	1.010	0.236	1.535
276	-0.302	-2.882	0.241	1.062	-0.041	-0.211
278	-0.311	-1.248	0.535	1.906	0.225	0.824
285	-0.450	-2.065	-0.044	-0.258	0.138	0.899
290	-0.594	-3.199	-0.752	-2.344	-0.766	-3.386
310	-0.493	-3.162	-0.289	-1.008	-0.365	-1.873
331	-0.305	-1.813	-0.846	-2.051	-0.149	-0.446
332	-0.055	-0.238	0.716	2.035	1.231	4.057
401	0.387	6.630	0.104	0.966	0.435	3.402
402	0.267	2.331	0.237	0.847	0.044	0.148
412	0.411	2.221	-0.121	-0.328	0.058	0.142
413	0.233	1.280	0.115	0.361	-0.366	-1.032
414	-0.166	-2.214	-0.073	-0.467	-0.061	-0.424
424	0.195	1.032	-0.314	-0.849	0.299	0.926
426	0.086	0.504	0.151	0.543	0.255	0.992
427	0.135	1.838	-0.027	-0.156	0.024	0.154
428	-0.112	-1.162	0.085	0.368	-0.239	-1.353
429	-0.287	-2.440	0.669	2.935	0.221	1.069
437	-0.146	-1.263	-1.172	-2.916	-0.277	-0.883
438	-0.075	-0.500	1.306	3.338	0.702	2.108
440	-0.172	-1.339	-0.065	-0.213	-0.025	-0.100
496	-0.007	-0.062	-0.003	-0.015	0.157	0.901
599	0.046	0.490	0.249	1.308	0.031	0.198
715	0.031	0.218	-0.032	-0.177	0.271	1.282
780	-0.352	-3.328	0.573	1.734	0.296	1.091
788	0.259	1.123	-0.191	-0.592	-0.152	-0.475

IV. SUGGESTIONS FOR FURTHER RESEARCH

The primary technical difficulty posed by the study of principal and secondary diagnoses is the overwhelming number of categories and possible combinations of categories. Rather than attempt to improve on the current research with even more elaborate statistical models, future work needs to focus on the use of medical expertise to reduce the complexity of the current models. Despite the use of "other" categories, and the relatively crude nature of the principal and secondary categorizations, the results from the present models still do not produce a manageable summary of this large database. More useful, and likely to be more reliable, sets of predicting equations could be formed using techniques similar to those used in this Note, but applied to a much coarser grouping of the secondary diagnoses. The number of such secondary diagnosis categories should be at most 10 or 15, and they must be formed to contain medically related conditions. (The results presented in this Note may be of some help in forming these categories, but the decisions should be dominated by clinical experience.)

There are several advantages to this approach over the current presentation. Aside from the fact that there would be many fewer estimates for a researcher to consider, the coarser groupings of secondary diagnoses would be common to each DRG (and most principal diagnoses), making the comparisons of secondary diagnoses for patients in different DRG categories much simpler than in the present work where many secondary diagnoses appear in several, but not all, of the DRG categories under study. In this Note, the most frequently occurring secondary diagnoses were determined separately for each DRG and used in the modeling of posthospital use. These selection criteria resulted in substantial undesirable variability in the secondary diagnoses that were modeled for each DRG.

Another advantage of coarser groupings of the secondary diagnoses is that it will permit more careful examination of possible interactions between the secondary and primary diagnoses, and may allow formal testing for differences in the effects of secondary diagnoses for patients in different DRG categories. This is likely to result in more reliable and medically plausible predicting models.

One primary weakness of the results presented in this Note is the questionable reliability of the recording of the secondary diagnoses for each patient. Although it was possible to check the validity of the models' fit in this Note by predicting posthospital use for patients in our database, this check was not very useful for determining whether some

of the differences among secondary diagnoses were simply record keeping anomalies. This approach would also fail to measure how much larger differences might be if accurate secondary information were used. Other data sources are needed to perform useful validity checks. Since they would not have to be used for reestimating models, other data sources can be much smaller than the MEDPAR files. One possible source of such data is the Kane study of posthospital use currently being conducted at the University of Minnesota. Another possible source is the DRG study currently being conducted at The RAND Corporation, although it may be more difficult to obtain necessary follow-up data after hospital discharge. There may also be difficulty identifying chronic conditions often indicated by secondary diagnoses from the hospital medical chart information collected for that study.

A more complete dataset, such as the one being collected at the University of Minnesota, may also be helpful when trying to interpret some of the estimated coefficients, especially the negative estimates for home health use. Obtaining more accurate information on the institutional status of patients using such care is much more desirable than attempting to explain such values by more complex and fragile statistical models, such as multinomial logit models that attempt to model each type of posthospital use simultaneously. Such an approach would be particularly difficult to apply in the posthospital utilizaton setting because numerous patterns of joint use are possible. For example, it is quite feasible for a patient to enter a REHAB facility, then to be discharged to further care in a SNF, and then to return home with follow-up care from an HHA.

Appendix A

RESULTS FOR THE OTHER DRG CATEGORIES

DRG 88: CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Patients are admitted into DRG 88 on the basis of their principal diagnosis. Most suffer from bronchitis, emphysema, or otherwise unclassified airway obstruction. As shown in Table A.1, almost 70 percent of our sample had a principal diagnosis of 49600, chronic airway obstruction, not elsewhere classified. This provided a natural point of reference for posthospital care use among the remainder of the patients, and it was used as the reference on which the principal parameter estimates were based. Principal diagnoses 50640 and 50690 were excluded from the regressions, as they have only three patients each, none of whom used posthospital care.

Table A.1

PRINCIPAL DIAGNOSES: DRG 88

Code	Description	Live Discharge	REHAB Use, %	SNF Use, %	HHA Use, %	No Use, %
49110	Mucopurulent chronic bronchitis	164	0.0	1.2	21.3	77.4
49120	Obstructive chronic bronchitis	4,775	0.1	1.1	15.7	83.2
49180	Other chronic bronchitis	124	0.0	0.0	11.3	88.7
49190	Unspecified chronic bronchitis	460	0.0	0.7	8.7	90.7
49200	Emphysematous bleb	213	0.0	1.9	18.3	79.8
49280	Other emphysema	3,652	0.2	1.6	18.4	79.8
49400	Bronchiectasis	958	0.1	0.6	14.5	84.8
49600	Chronic airway obstruction nec	24,238	0.1	1.5	18.0	80.4
50640	Chronic respiratory conditions due to fumes and vapors	3	0.0	0.0	0.0	100.0
50690	Unspecified respiratory conditions due to fumes and vapors	3	0.0	0.0	0.0	100.0
	Total	34,595	0.1	1.4	17.5	81.0

DRG 88 differs from our other DRG categories in that its patients make relatively little use of any type of posthospital care. Only 19 percent use posthospital care, as opposed to about 57, 54, 41, and 22 percent for DRG categories 14, 209, 210, and 127, respectively. Home health care accounts for the vast majority of all posthospital care used by DRG 88 patients. On average, 17.5 percent of patients use HHA care. Only 1.4 percent use SNF care, and 0.1 percent use REHAB care. The fact

that so few DRG 88 patients use REHAB care means that the regression techniques that we used for DRG 14 cannot be used to distinguish differences in REHAB use between patients with different secondary diagnoses. Even if a patient's chances of needing REHAB care increase from 0.1 percent to 0.2 percent, which represents a doubling of the odds of REHAB use, this is not a particularly relevant increase for our purposes. Such changes, if they exist, cannot be detected with our statistical methods, even with the large sample sizes available to us. As a consequence, we present analyses for HHA and SNF care only.

Table A.2 shows the percentages of patients using posthospital care for each secondary diagnosis. Diagnoses assigned to at least 350 patients, plus a few assigned to at least 300 patients, are included. Home health care use varies from 12.8 percent for patients with diverticula of the intestine (562), to 28.6 percent for those with bone and cartilage disorders (733). SNF care ranges between 0.3 percent for patients suffering from obesity and other hyperalimentation (278), and 4.3 percent for patients with specific nonpsychotic mental disorders resulting from organic brain damage (310). As with Table 2.1 for DRG 14, some dependence between secondary and principal diagnoses is evident, but far less than would be needed to explain most of the variation among secondary diagnoses. The secondary diagnoses with the most extreme values of posthospital care are evenly distributed among the different principal diagnoses.

Examining Table A.3, other trends among secondary diagnoses are evident. There is a strong positive correlation between HHA and SNF care; secondary diagnosis categories whose patients use much of one type of care tend also to use the other type frequently. Patients suffering from heart ailments, bone and cartilage disorders, anemias, and urinary tract problems used relatively more posthospital care; those suffering from obesity or with respiratory problems such as asthma, emphysema, and bronchitis used relatively little. The regression analyses confirmed these differences even after the adjustment for differences in principal diagnosis. Of the diagnoses relating to heart problems, 416, 427, and 428 all have statistically significant increases in a patient's chance of using both HHA and SNF care. This finding is particularly strong for diagnosis 428, heart failure, which has a t-statistic of 10.17 and an estimate of 0.45 for HHA care use. Bone and cartilage disorders (733), and other disorders of the urethra and urinary tract (599), also had strong positive effects on both HHA and

Table A.2

SECONDARY DIAGNOSES: DRG 88

Code	Description	Live Discharge	REHAB. Use, %	SNF Use, %	HHA Use, %	No Use, %
V10	Personal history of malignant neoplasm	756	0.4	0.4	16.7	82.5
V12	Personal history of certain other diseases	368	0.0	1.1	17.4	81.5
V45	Other postsurgical states	366	0.0	0.8	13.4	85.8
041	Bacterial infection in conditions classified elsewhere and of unspecified site	1,268	0.1	2.0	18.3	79.7
162	Malignant neoplasm of trachea, bronchus, and lung	373	0.3	1.1	23.6	75.1
244	Acquired hypothyroidism	343	0.3	1.7	21.0	77.0
250	Diabetes mellitus	2,163	0.0	1.3	19.6	79.1
276	Disorders of fluid, electrolyte, and acid-base balance	2,400	0.1	2.9	21.4	75.7
278	Obesity and other hyperalimentation	385	0.0	0.3	13.2	86.5
285	Other and unspecified anemias	418	0.0	2.9	24.9	72.2
300	Neurotic disorders	919	0.3	1.5	17.6	80.5
310	Specific nonpsychotic mental disorders due to organic brain damage	376	0.0	4.3	18.6	77.1
311	Depressive disorder, nec	445	0.2	1.8	21.6	76.4
401	Essential hypertension	2,881	0.2	1.0	16.3	82.5
402	Hypertensive heart disease	539	0.2	0.9	17.1	81.8
412	Old myocardial infarction	505	0.0	0.4	14.9	84.8
413	Angina pectoris	1,201	0.0	0.6	17.8	81.6
414	Other forms of chronic ischemic heart disease	4,442	0.0	1.4	17.6	81.0
416	Chronic pulmonary heart disease	2,275	0.3	2.3	22.6	74.8
426	Conduction disorders	551	0.2	0.4	16.7	82.8
427	Cardiac dysrhythmias	3,516	0.2	2.2	21.2	76.5
428	Heart failure	4,942	0.1	2.1	23.5	74.3
429	Ill-defined descriptions and complications of heart disease	1,262	0.3	0.7	17.8	81.1
440	Atherosclerosis	569	0.2	1.1	20.2	78.6
465	Acute upper respiratory infections of multiple or unspecified sites	460	0.2	1.3	16.3	82.2
466	Acute bronchitis and bronchiolitis	2,893	0.1	1.2	16.6	82.1
486	Pneumonia, organism unspecified	1,079	0.0	2.6	18.9	78.5
490	Bronchitis, not specified as acute or chronic	622	0.3	1.8	14.0	83.9
491	Chronic bronchitis	798	0.1	1.1	17.0	81.7
492	Emphysema	688	0.0	1.7	14.4	83.9
493	Asthma	1,448	0.1	0.8	14.4	84.7
496	Chronic airway obstruction, nec	755	0.1	1.5	18.0	80.4
515	Postinflammatory pulmonary fibrosis	544	0.4	1.5	14.7	83.5
518	Other diseases of lung	825	0.2	1.7	17.7	80.4
519	Other diseases of respiratory system	1,580	0.1	0.9	14.5	84.6
530	Diseases of esophagus	485	0.4	0.8	15.3	83.5
535	Gastritis and duodenitis	562	0.0	1.2	18.7	80.1
553	Other hernia of abdominal cavity without mention of obstruction or gangrene	502	0.0	1.0	13.9	85.1
562	Diverticula of intestine	366	0.3	1.4	12.8	85.5
599	Other disorders of urethra & urinary tract	987	0.0	3.3	24.6	72.0
715	Osteoarthrosis and allied disorders	858	0.0	1.3	18.9	79.8
733	Other disorders of bone and cartilage	986	0.1	2.6	28.6	68.7
780	General symptoms	672	0.3	3.0	18.3	78.4
786	Symptoms involving respiratory system and other chest symptoms	4,299	0.2	1.3	16.1	82.4
799	Other ill-defined and unknown causes of morbidity and mortality	3,617	0.1	2.4	22.2	75.3
	Total	34,595	0.1	1.4	17.5	81.0

Table A.3

ESTIMATED MODELS FOR DRG 88

Code	HHA Est.	HHA t-stat.	SNF Est.	SNF t-stat.
		Primary Diagnoses		
49110	0.27	1.32	-0.10	-0.13
49120	-0.14	-3.07	-0.32	-2.03
49180	-0.43	-1.44	-5.98	-0.65
49190	-0.78	-4.37	-0.39	-0.78
49200	0.03	0.14	0.02	0.03
49280	0.02	0.36	-0.01	-0.04
49400	-0.20	-2.06	-0.89	-1.96
		Secondary Diagnoses		
V10	—	—	-1.05	-1.81
162	0.51	3.80	—	—
250	0.19	2.96	—	—
276	0.30	4.88	0.69	4.44
278	-0.37	-2.24	—	—
285	0.54	4.35	—	—
310	—	—	1.12	3.99
413	—	—	-0.92	-2.23
416	0.26	4.38	0.56	3.47
426	—	—	-1.24	-1.75
427	0.25	4.63	0.49	3.27
428	0.45	10.17	0.44	3.39
486	—	—	0.44	1.93
493	-0.14	-1.69	—	—
519	-0.21	-2.57	—	—
599	0.45	5.42	0.90	4.47
733	0.71	8.31	0.78	3.32
780	—	—	0.59	2.28
799	0.34	6.86	0.57	4.14
Other	0.16	3.76	0.36	2.48
Interact	-0.08	-1.57	-0.13	-0.87

SNF use. Anemias (285), while having a measurable effect on HHA care use, did not appear to be predictive of SNF use.

The chi square tests for the hypothesis that secondary diagnoses carry no additional information about HHA or SNF use were highly significant with chi square values of 379 and 282. A chi square value larger than 75 is extremely unlikely.

DRG 127: HEART FAILURE AND SHOCK

Heart failure and shock patients are admitted into DRG 127 on the basis of their principal diagnosis. The vast majority, 85 percent, suffer from congestive heart failure, as shown in Table A.4. This principal diagnosis was therefore used as the

Table A.4

PRINCIPAL DIAGNOSES: DRG 127

Code	Description	Live Discharge	REHAB Use, %	SNF Use, %	HHA Use, %	No Use, %
39891	Rheumatic congestive heart failure	323	0.0	1.5	23.5	74.9
40201	Malignant hypertensive heart disease with congestive heart failure	186	0.0	0.5	23.1	76.3
40211	Benign hypertensive heart disease with congestive heart failure	590	0.0	1.7	19.2	79.2
40291	Unspecified hypertensive heart disease with congestive heart failure	4,094	0.1	1.8	20.3	77.7
42800	Congestive heart failure	78,141	0.1	2.1	20.2	77.5
42810	Left heart failure	7,067	0.0	2.1	18.2	79.7
42890	Unspecified heart failure	456	0.0	3.5	19.3	77.2
78550	Unspecified shock without mention of trauma	87	1.1	4.6	14.9	79.3
78551	Cardiogenic shock without mention of trauma	121	0.0	2.5	9.9	87.6
78559	Other shock without mention of trauma	515	0.6	7.6	14.0	77.9
	Total	91,588	0.1	2.1	20.0	77.7

reference for comparing the other principal diagnoses. The broader three-digit category of heart failure (428) describes 93 percent of DRG 127 patients. Since there is very little variation in the principal diagnosis among the DRG 127 patients, the differences in posthospital use among patients with different secondary diagnoses, shown in Table A.5, can be more safely attributed to the secondary diagnoses than to the dependence of secondary diagnoses on the principal diagnosis. DRG 127 patients make relatively little use of any type of posthospital care compared to all the other DRG categories studied except DRG 88. Home health care is most frequently used, averaging 20 percent, with a range of 9.9 percent for cardiogenic shock patients to 23.5 percent for rheumatic congestive heart failure patients. SNF care use averages only 2.1 percent and REHAB care only 0.1 percent. Because there is so little REHAB use, analyses were not performed for this type of posthospital care.

Table A.5 shows posthospital care use for each secondary diagnosis assigned to at least 1,000 live discharges, plus a few assigned to at least 800. Home health care ranges from 16.9 percent for patients with old myocardial infarctions (412), to 27.2

Table A.5

SECONDARY DIAGNOSES: DRG 127

	Description	Live Discharge	REHAB Use, %	SNF Use, %	HHA Use, %	No Use, %
V10	Personal hist. of malignant neoplasm	1,384	0.0	1.8	20.0	78.2
V45	Other postsurgical states	3,510	0.1	1.6	17.8	80.5
041	Bacterial infection in conditions classified elsewhere and of unspecified site	2,052	0.1	5.2	25.9	68.9
244	Acquired hypothyroidism	1,242	0.1	2.5	22.5	75.0
250	Diabetes mellitus	16,916	0.1	1.7	21.7	76.4
276	Disorders of fluid, electrolyte, and acid-base balance	5,883	0.1	4.0	21.7	74.2
278	Obesity & other hyperalimentation	1,264	0.1	0.6	17.2	82.1
280	Iron deficiency anemias	1,930	0.0	2.2	22.1	75.8
285	Other & unspecified anemias	3,266	0.0	2.8	22.2	75.0
310	Specific nonpsychotic mental disorders due to organic brain damage	1,369	0.1	6.3	21.7	72.0
401	Essential hypertension	9,492	0.1	1.4	19.0	79.5
402	Hypertensive heart disease	2,655	0.0	0.9	17.7	81.4
403	Hypertensive renal disease	897	0.0	3.0	22.3	74.7
411	Other acute and subacute forms of ischemic heart disease	3,231	0.0	1.0	20.8	78.1
412	Old myocardial infarction	3,773	0.1	1.1	16.9	81.9
413	Angina pectoris	5,395	0.0	1.3	18.3	80.4
414	Other forms of chronic ischemic heart disease	28,111	0.1	1.9	19.0	79.0
416	Chronic pulmonary heart disease	996	0.0	1.4	22.8	75.8
424	Other diseases of endocardium	5,297	0.0	1.9	19.3	78.8
425	Cardiomyopathy	5,473	0.1	1.6	20.4	78.0
426	Conduction disorders	4,518	0.2	1.7	18.1	80.0
427	Cardiac dysrhythmias	25,517	0.1	2.2	20.2	77.5
428	Heart failure	882	0.1	2.4	19.2	78.3
429	Ill-defined descriptions & complications of heart disease	7,869	0.1	1.9	18.9	79.1
437	Other and ill-defined cerebrovascular disease	1,226	0.3	3.9	21.4	74.4
438	Late effects of cerebrovascular disease	1,144	0.8	4.5	26.4	68.4
440	Atherosclerosis	2,836	0.2	2.2	19.0	78.6
443	Other peripheral vascular disease	1,231	0.1	2.5	22.3	75.1
466	Acute bronchitis and bronchiolitis	2,112	0.0	2.5	18.5	78.9
491	Chronic bronchitis	1,319	0.1	1.8	18.9	79.2
492	Emphysema	1,305	0.0	1.8	19.2	79.0
493	Asthma	1,162	0.0	1.9	18.8	79.3
496	Chronic airway obstruction, nec	11,776	0.1	2.0	20.6	77.3
511	Pleurisy	6,382	0.2	3.3	23.3	73.2
514	Pulmonary congestion and hypostasis	1,030	0.2	2.5	22.4	74.9
518	Other diseases of lung	1,099	0.2	2.6	20.9	76.3
585	Chronic renal failure	2,734	0.0	2.9	19.1	77.9
586	Renal failure, unspecified	1,264	0.1	4.1	23.6	72.2
593	Other disorders of kidney and ureter	1,807	0.1	2.7	22.2	75.0
599	Other disorders of urethra and urinary tract	5,406	0.2	5.0	25.8	69.0
715	Osteoarthrosis and allied disorders	3,110	0.1	2.3	21.9	75.7
780	General symptoms	1,257	0.3	2.5	21.0	76.2
782	Symptoms involving skin and other integumentary tissue	1,004	0.0	2.9	27.2	69.9
786	Symptoms involving respiratory system and other chest symptoms	1,877	0.1	1.4	19.7	78.8
789	Other symptoms involving abdomen and pelvis	1,119	0.2	2.3	22.3	75.2
790	Nonspecific findings on examination of blood	1,404	0.1	2.6	21.2	76.2
799	Other ill-defined and unknown causes of morbidity and mortality	1,844	0.4	3.9	24.7	71.0
	Total	91,588	0.1	2.1	20.0	77.7

percent for those with skin diseases (782). An analysis of the relationship between principal and secondary diagnoses, like that given in Table 2.1 for DRG 14, shows that secondary codes 428, heart failure, and 599, urinary tract disorders, are frequently assigned to shock patients, who use the least HHA care. This dependence explains some of the association between the secondary diagnoses and posthospital use, but appears to be an inadequate explanation for the differences in most of the secondary diagnoses. To obtain a more formal assessment of this statement, standard regression analyses were used.

The regression results are summarized in Table A.6. Secondary diagnosis 428 does not have much predictive value for HHA aside from the principal diagnoses associated with it and is not included among the parameters with significant t-statistics. Diagnosis 599, on the other hand, rather than being explained by the effect of principal diagnoses, acts very powerfully and to some extent counters the principal diagnoses' negative effect on HHA care use. Urinary tract disorders have the most significant and one of the largest effects on HHA care use of the 13 secondaries we measured. Its t-statistic is 6.85, with an estimate of 0.41. Pleurisy, secondary diagnosis 511, has the largest t-statistic for HHA care use and also has a positive effect.

DRG 209: MAJOR JOINT AND LIMB REATTACHMENT PROCEDURES

Of DRG 209 patients, 28.6 percent use HHA care, more than in any of the other four DRG categories analyzed. In terms of SNF and REHAB care, they fall in the middle range: 15.2 percent and 2.2 percent, respectively, use these types of care. Unlike the other DRG categories we have analyzed, DRG 209 is procedure-driven. It is assigned not according to principal diagnosis but according to whether at least one of a number of procedures has been performed on the patient. For this reason many principal diagnoses are represented among DRG 209 patients. Nevertheless, some are much more common than others, and they are listed in Table A.7. Those not listed constitute the "other" category, and they are used as the reference for principal diagnoses in the logistic regressions. Principal diagnoses may be conveniently grouped into fracture and nonfracture cases. Patients with fractures tend to use more REHAB and SNF care, and less HHA care.

Table A.6

ESTIMATED MODELS FOR DRG 127

	HHA		SNF	
Code	Est.	t-stat.	Est.	t-stat.
Primary Diagnoses				
39891	0.38	1.62	-0.11	-0.15
40201	0.20	0.65	-0.35	-0.34
40211	-0.24	-1.22	0.02	0.03
40291	0.10	1.41	-0.48	-2.05
42810	-0.13	-2.19	0.09	0.60
42890	-0.19	-0.84	0.65	1.53
78550	-0.66	-1.24	1.17	1.90
78551	-2.30	-2.30	1.17	1.90
78559	-0.41	-1.93	-5.56	-0.60
Secondary Diagnoses				
250	0.17	3.97	—	—
276	0.21	3.37	0.84	6.62
278	—	—	-2.03	-2.02
310	—	—	1.13	5.56
402	—	—	-0.69	-2.04
411	0.17	2.11	-0.71	-2.19
413	—	—	-0.62	-2.56
414	-0.05	-1.27	—	—
426	-0.12	-1.59	—	—
427	—	—	0.30	3.38
437	—	—	0.59	2.11
438	0.32	2.64	0.98	4.00
443	—	—	0.75	2.97
496	0.13	2.82	—	—
511	0.36	6.49	0.60	4.53
514	0.28	2.13	—	—
586	0.26	2.31	—	—
599	0.41	6.85	0.96	7.78
782	0.43	3.36	—	—
799	0.37	3.66	0.71	3.22
Other	0.13	2.66	0.49	3.09
Interact	-0.10	-2.01	-0.26	-2.52

For each secondary diagnosis, the propensity to use posthospital care is given in Table A.8. Home health care use ranges from a low of 13.2 percent for patients with senile and presenile organic psychotic conditions (290), to 36.6 percent of patients with angina pectoris (413). Patients with mental problems tend to have low HHA care use, whereas those with various types of heart disease use slightly more than average.

The analysis of the relationship between principal and secondary diagnoses shows stronger and more easily interpretable correlations in this DRG. For example, accidental falls, secondary diagnosis E88, typically resulted in classification into one of the fracture categories. All four of the mental disorders (290, 310, 331, and 332) are also correlated with fracture cases. Even when adjustment is made to account for

Table A.7

PRINCIPAL DIAGNOSES: DRG 209

Code	Description	Live Disch.	REHAB. Use, %	SNF Use, %	HHA Use, %	No Use, %
71515	Localized osteoarthropathy, idiopathic, pelvic region and thigh	553	0.90	8.50	28.57	62.03
71516	Localized osteoarthropathy, idiopathic, lower leg	705	1.99	4.40	32.34	61.28
71535	Osteoarthrosis, localized, not specific whether principal or secondary, pelvic region and thigh	1,700	1.18	7.12	25.59	66.12
71536	Osteoarthrosis, localized, not specific whether principal or secondary, lower leg	2,056	2.33	5.70	28.21	63.76
71595	Osteoarthrosis, unspecified whether generalized or localized, pelvic region and thigh	4,128	1.62	6.40	29.46	62.52
71596	Osteoarthrosis, unspecified whether generalized or localized, lower leg	4,700	1.62	6.45	31.55	60.38
73342	Aseptic necrosis of bone, head and neck of femur	698	2.01	9.60	29.66	58.74
82000	Transcervical fracture, closed, intracapsular section, unspecified	741	1.89	30.50	25.51	42.11
82002	Closed, transcervical fracture of neck of femur, midcervical section	502	3.98	30.88	24.90	40.24
82009	Transcervical fracture, closed, midcervical section	5,464	2.75	29.36	26.87	41.03
82021	Pertrochanteric fracture, closed, intertrochanteric section	452	3.10	35.62	20.13	41.15
	Total	30,889	2.20	15.18	28.59	54.02

the tendency of these secondary diagnoses to occur with a particular class of principal diagnosis, the secondary diagnoses still appeared to be predictive of posthospital use based on the regression analyses.

Table A.8

SECONDARY DIAGNOSES: DRG 209

Code	Description	Live Discharge	REHAB Use, %	SNF Use, %	HHA Use, %	No Use, %
E88	Accidental falls	1,323	3.3	28.6	25.9	42.2
V10	Personal history of malignant neoplasm	648	2.6	14.5	30.7	52.2
V43	Organ or tissue replaced by other means	597	3.2	7.0	27.8	62.0
V45	Other postsurgical states	345	2.9	14.8	27.8	54.5
041	Bacterial infection in conditions classified elsewhere and of unspecified site	1,210	2.6	19.8	29.4	48.2
244	Acquired hypothyroidism	440	1.8	14.8	26.8	56.6
250	Diabetes mellitus	1,848	2.8	17.8	30.6	48.8
276	Disorders of fluid, electrolyte, and acid-base balance	1,007	3.0	25.4	27.5	44.1
278	Obesity and other hyperalimentation	610	3.3	10.7	28.2	57.9
280	Iron deficiency anemias	457	2.6	17.3	29.8	57.9
285	Other and unspecified anemias	1,705	2.2	18.2	30.1	49.5
290	Senile and presenile organic psychotic conditions	468	1.1	35.3	13.2	50.4
310	Specific nonpsychotic mental disorders due to organic brain damage	586	2.0	35.3	13.1	49.5
331	Other cerebral degenerations	279	0.7	40.1	15.1	44.1
332	Parkinson's disease	379	4.0	33.8	19.5	42.7
401	Essential hypertension	4,923	2.4	12.6	30.5	54.5
402	Hypertensive heart disease	603	2.8	19.4	29.9	47.9
412	Old myocardial infarction	346	2.0	14.2	28.0	55.8
413	Angina pectoris	473	2.5	13.7	36.6	47.1
414	Other forms of chronic ischemic heart disease	2,171	2.2	22.8	28.7	46.3
424	Other diseases of endocardium	346	2.0	16.2	31.2	50.6
426	Conduction disorders	650	2.5	16.2	24.6	56.8
427	Cardiac dysrhythmias	1,806	2.1	22.5	29.2	46.2
428	Heart failure	926	2.7	28.5	27.5	41.3
429	Ill-defined descriptions and complications of heart disease	971	3.7	19.5	28.1	48.7
437	Other and ill-defined cerebrovascular disease	293	0.0	32.1	19.1	48.8
438	Late effects of cerebrovascular disease	306	4.9	27.1	27.1	40.8
440	Atherosclerosis	523	2.5	22.8	22.9	51.8
496	Chronic airway obstruction, nec	1,150	2.3	20.5	29.0	48.2
518	Other diseases of lung	309	3.2	22.3	28.2	46.3
553	Other hernia of abdominal cavity without mention of obstruction or gangrene	341	1.2	14.4	31.7	52.8
599	Other disorders of urethra and urinary tract	2,101	2.8	22.1	30.4	44.7
600	Hyperplasia of prostate	309	0.6	12.3	22.7	64.4
714	Rheumatoid arthritis and other inflammatory polyarthropathies	555	2.9	15.0	30.8	51.4
715	Osteoarthrosis and allied disorders	1,598	2.3	17.1	30.5	50.1
718	Other derangement of joint	336	2.4	9.2	36.3	52.1
733	Other disorders of bone & cartilage	1,237	2.3	20.8	30.3	46.6
736	Other acquired deformities of limbs	520	1.9	9.2	25.6	63.3
780	General symptoms	434	3.7	17.7	29.7	48.8
788	Symptoms involving urinary system	470	1.7	19.8	31.7	46.8
996	Complications peculiar to certain specified procedures	618	2.9	16.2	30.6	50.3
997	Complications affecting specified body systems, nec	965	3.0	15.6	30.7	50.7
998	Other complications of procs., nec	1,119	2.3	13.8	31.4	52.5
	Total	123,556	2.2	15.2	28.6	54.0

Although the use of HHA decreases in the presence of mental disorders, SNF use increases. Among the secondary codes in Table A.8, the highest SNF use is among patients with secondary code 331, other cerebral degenerations. Over 40 percent of these patients use SNF care. The three other categories of mental disorders also have high propensities to use SNF care. The regression results show that all four types have strong and highly significant positive effects on SNF care use. Patterns of SNF use also differ from HHA care in their response to heart troubles. SNF use varies significantly among a number of heart-related secondary diagnoses, although the direction is not entirely consistent. Table A.8 shows that patients with essential hypertension and hypertensive heart disease (401 and 402) use less SNF care, and those with chronic ischemic heart disease, cardiac dysrythmias, and heart failure (414, 427, and 428, respectively) all use more care. In each case this is borne out by the regression results shown in Table A.9.

Only a few secondary diagnoses significantly differ in the propensity of DRG 209 patients to use REHAB care. As with HHA care and SNF care use, patients with mental disorders tend to use less REHAB care. Patients with late effects of cerebrovascular disease (438) use more REHAB care than those assigned any other secondary code. Obesity (278) and ill-defined descriptions and complications of heart disease (429) also have positive indications.

The chi square tests for the hypothesis that secondary diagnoses carry no additional information about HHA or SNF use were highly significant with chi square values of 262 and 326. A chi square value larger than 75 is extremely unlikely. The chi square statistic for REHAB, 61, is only marginally significant at about the 0.05 level.

DRG 210: HIP AND FEMUR PROCEDURES EXCEPT MAJOR JOINT, AGE ≥ 70 AND/OR COMPLICATIONS

Like DRG 209, patients qualify for inclusion in DRG 210 based on the procedures performed on them. REHAB care is used by 2.7 percent of DRG 210 patients, SNF care is used by 30.5 percent, and HHA care by 25.6 percent. DRG 210 has several unique characteristics. These patients use far more SNF care than those in any other DRG, and this is the only DRG in which SNF care is actually used more frequently than HHA care. As Table A.10 shows, there is relatively little variation in posthospital care use among patients with different principal diagnoses. The analysis of the dependence of secondary diagnoses on principal diagnosis also indicated very

Table A.9

ESTIMATED MODELS FOR DRG 209

	HHA		SNF		REHAB	
Code	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
71515	-0.19	-2.06	-0.74	-4.74	-1.08	-2.38
71516	-0.06	-0.70	-1.44	-7.75	-0.23	-0.87
71535	-0.30	-5.20	-0.92	-9.35	-0.82	-3.47
71536	-0.17	-3.14	-1.14	-11.62	-0.11	-0.72
71595	-0.14	-3.40	-1.03	-14.77	-0.44	-3.18
71596	-0.03	-0.76	-1.03	-15.59	-0.47	-3.56
73342	-0.08	-0.96	-0.61	-4.67	-0.28	-1.02
82000	0.01	0.13	0.63	7.35	-0.30	-1.13
82002	0.01	0.09	0.67	6.64	0.41	1.72
82009	0.09	2.46	0.57	13.79	0.02	0.20
82021	-0.16	-1.55	0.84	8.18	0.13	0.45
E88	—	—	0.21	3.25	0.20	1.23
V43	—	—	-0.60	-3.66	0.44	1.86
250	0.13	2.49	—	—	—	—
276	—	—	0.32	3.92	—	—
278	—	—	—	—	0.48	2.05
285	0.15	2.75	—	—	—	—
290	-0.94	-7.60	0.41	3.91	-1.15	-2.28
310	-0.90	-8.29	0.52	5.55	—	—
331	-0.67	-4.47	0.71	5.48	-1.30	-1.82
332	—	—	0.60	5.20	0.50	1.85
401	0.07	1.84	-0.17	-3.14	—	—
413	—	—	-0.31	-2.18	—	—
414	—	—	0.30	5.09	—	—
427	—	—	0.30	4.64	—	—
428	—	—	0.36	4.55	—	—
429	—	—	—	—	0.47	2.65
437	-0.45	-3.21	—	—	—	—
438	—	—	—	—	0.79	3.09
553	—	—	—	—	-0.70	-1.38
599	0.12	2.28	0.12	1.92	—	—
600	-0.47	-3.50	—	—	-1.25	-1.76
718	0.28	2.40	—	—	—	—
733	0.15	2.37	—	—	—	—
736	-0.13	-1.32	—	—	—	—
780	—	—	—	—	0.42	1.64
788	0.20	2.02	—	—	—	—
997	—	—	—	—	0.33	1.70
Other	0.05	1.52	0.12	2.88	0.24	2.67
Interact	0.04	1.02	0.13	2.73	—	—

Table A.10

PRINCIPAL DIAGNOSES: DRG 210

Code	Description	Live Discharge	REHAB Use, %	SNF Use, %	HHA Use, %	No Use, %
82000	Closed transcervical fracture of intracapsular section of neck of femur, unspecified	235	2.1	25.1	27.2	45.5
82002	Closed transcervical fracture of midcervical section of neck of femur	238	2.5	23.1	29.8	44.5
82003	Closed transcervical fracture of base of neck of femur	549	3.1	35.2	24.0	37.7
82009	Other closed transcervical fracture of neck of femur	2,174	2.1	26.3	27.1	44.5
82020	Closed pertrochanteric fracture of the trochanteric section of neck of femur, unspecified	287	3.1	34.5	21.3	41.1
82021	Closed pertrochanteric fracture of the intertrochanteric section of neck of femur	13,309	2.8	33.6	24.2	39.4
82022	Closed pertrochanteric fracture of the subtrochanteric section of neck of femur	870	3.2	31.6	26.2	39.0
82031	Open pertrochanteric fracture of the intertrochanteric section of neck of femur	245	5.7	28.6	26.5	39.2
82100	Closed fracture of other and unspecified part of femur, thigh or upper leg	189	1.1	25.9	39.2	33.9
82101	Closed fracture of other and unspecified part of shaft of femur	426	3.5	28.2	24.4	43.9
82123	Closed supracondylar fracture of other and unspecified part of lower end of femur	326	1.8	26.4	27.9	43.9
	Total	22,942	2.7	30.5	25.6	41.3

little dependence. It is expected, therefore, that the variation in posthospital care use among the secondary diagnoses shown in Table A.11 cannot be explained by variation in principal diagnoses.

Examining Table A.11, we detect trends in posthospital care use among patients with different secondary diagnoses. REHAB care varies between a low of 0.2 percent for patients with personal histories of malignant neoplasms (V10), and a high of 5.7 percent for patients with Parkinson's disease. This high rate for Parkinson's disease is

Table A.11

SECONDARY DIAGNOSES: DRG 210

Code	Description	Live Discharge	REHAB Use,%	SNF Use,%	HHA Use,%	No Use,%
E88	Accidental falls	2,547	2.6	30.0	24.1	43.3
V10	Personal history of malignant neoplasm	552	0.2	29.0	29.9	38.2
V12	Personal history of certain other diseases	207	4.3	25.6	23.7	46.4
V45	Other postsurgical states	250	2.8	31.6	28.4	37.2
041	Bacterial infection in conditions classified elsewhere and of unspecified site	1,253	2.0	31.9	24.8	41.3
198	Secondary malignant neoplasm of other specified sites	334	1.2	20.4	32.6	45.8
244	Acquired hypothyroidism	245	3.7	33.1	26.5	36.7
250	Diabetes mellitus	1,770	3.3	30.7	27.4	38.6
276	Disorders of fluid, electrolyte, and acid-base balance	1,182	2.5	34.3	23.4	39.8
280	Iron deficiency anemias	520	3.7	33.8	24.8	37.7
281	Other deficiency anemia	233	2.6	39.5	25.8	32.2
285	Other and unspecified anemias	1,762	3.0	34.1	24.2	38.8
290	Senile and presenile organic psychotic conditions	772	0.8	36.8	9.8	52.6
298	Other nonorganic psychoses	244	0.4	30.3	14.8	54.5
310	Specific nonpsychotic mental disorders due to organic brain damage	1,054	1.5	33.9	11.0	53.6
331	Other cerebral degenerations	345	2.0	35.1	11.3	51.6
332	Parkinson's disease	418	5.7	34.2	17.9	42.1
342	Hemiplegia	245	4.9	32.7	23.7	38.8
401	Essential hypertension	2,645	3.8	30.1	28.6	37.5
402	Hypertensive heart disease	442	2.9	36.2	27.1	33.7
410	Acute myocardial infarction	143	1.4	38.5	20.3	39.9
412	Old myocardial infarction	225	2.7	31.1	25.3	40.9
413	Angina pectoris	308	1.9	36.0	25.3	36.7
414	Other forms of chronic ischemic heart disease	2,143	2.6	34.3	23.1	40.0
415	Acute pulmonary heart disease	141	2.1	30.5	34.8	32.6
424	Other diseases of endocardium	375	3.5	32.0	25.3	39.2
426	Conduction disorders	639	3.6	29.3	27.7	39.4
427	Cardiac dysrhythmias	1,856	2.6	37.6	22.0	37.8
428	Heart failure	1,325	2.1	34.7	21.2	42.0
429	Ill-defined descriptions and complications of heart disease	999	3.4	34.3	20.4	41.8
437	Other and ill-defined cerebrovascular disease	395	2.0	32.7	15.4	49.9
438	Late effects of cerebrovascular disease	425	4.9	33.2	22.8	39.1
440	Atherosclerosis	668	2.7	32.8	20.4	44.2
486	Pneumonia, organism unspecified	294	2.0	33.3	18.4	46.3
491	Chronic bronchitis	186	3.2	31.7	26.9	38.2
492	Emphysema	241	1.2	30.7	28.6	39.4
496	Chronic airway obstruction, nec	1,324	3.2	29.8	27.5	39.5
518	Other diseases of lung	269	1.1	36.1	23.8	39.0
553	Other hernia of abdominal cavity without mention of obstruction or gangrene	188	4.3	28.7	29.3	37.8
560	Intestinal obstruction without mention of hernia	234	1.7	34.6	21.8	41.9
599	Other disorders of urethra and urinary tract	2,556	2.5	33.2	23.7	40.5

Table 11 (continued)

Code	Description	Live Discharge	REHAB Use,%	SNF Use,%	HHA Use,%	No Use,%
707	Chronic ulcer of skin	318	2.2	36.5	19.2	42.1
714	Rheumatoid arthritis and other inflammatory polyarthropathies	197	5.1	24.4	32.0	38.6
715	Osteoarthrosis and allied disorders	904	3.4	32.0	25.6	39.0
733	Other disorders of bone and cartilage	1,153	3.1	30.8	21.9	44.1
780	General symptoms	528	4.0	32.0	21.8	42.2
788	Symptoms involving urinary system	379	2.4	31.1	28.2	38.3
797	Senility w/out mention of psychosis	198	0.5	34.3	13.1	52.0
799	Other ill-defined and unknown causes of morbidity and mortality	141	3.5	29.8	24.8	41.8
812	Fracture of humerus	387	3.9	43.7	18.3	34.1
813	Fracture of radius and ulna	436	4.6	32.3	22.5	40.6
820	Fracture of neck of femur	634	3.3	33.6	25.7	37.4
996	Complications peculiar to certain specified procedures	306	2.0	30.7	19.9	47.4
997	Complications affecting specified body systems, nec	578	3.5	34.1	24.2	38.2
998	Other complications of procedures, nec	588	2.2	33.0	22.4	42.3
	Total	91,768	2.7	30.5	25.6	41.3

unique among mental disorders. All of the other diagnoses relating to mental problems, 290, 298, 310, 331, and 797, have low rates of REHAB care use. Consistent with our findings in other DRG categories, patients with any type of mental disorder tend to use less HHA care and more SNF care. In fact, HHA care use is least common for patients with senile and presenile organic psychotic conditions (290). Only 9.8 percent of them use HHA care, as opposed to a high of 34.8 percent of patients with acute pulmonary heart disease (415).

Table A.12 shows the results of the logistic regressions. The parameter estimates are generally consistent with the size and direction of the percentages given in Table A.11. Acute pulmonary heart disease (415), however, whose patients have the highest HHA care use of any secondary diagnosis group, does not have a statistically significant estimate for HHA use after our crude adjustment for differences between principal diagnoses. The estimate is positive, but its t-statistic is low. This is likely to be due to the small sample size; only 141 patients had this secondary diagnosis. Patients with secondary diagnosis 812, fracture of humerus, are heavy users of SNF care. The other two fractures, however, do not appear to be strong predictors.

Only four of the secondary diagnoses we included in the final regression fit have a measurable difference in REHAB use. Two of these are mental disorders:

Parkinson's disease makes it much more likely that a patient will use REHAB care, whereas senile and presenile organic psychotic conditions (290) make it less likely. This is consistent with the general tendencies to use REHAB care found in Table A.11.

Although patients with all three fractures use more REHAB care, only fracture of radius and ulna (813) was included in the final regression. Table A.12 also shows that these patients are prone to greater use of REHAB care.

The chi square tests for the hypothesis that secondary diagnoses carry no additional information about HHA, SNF, or REHAB use were highly significant with chi square values of 625, 136, and 89. A chi square value larger than 75 is extremely unlikely.

Table A.12

ESTIMATED MODELS FOR DRG 210

	HHA		SNF		REHAB	
Code	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
Primary Diagnoses						
82000	0.03	0.24	0.11	0.71	-0.19	-0.41
82002	0.14	1.02	-0.01	-0.06	-0.04	-0.10
82003	0.05	0.48	0.60	6.17	0.24	0.93
82009	0.02	0.40	0.17	2.77	-0.18	-1.02
82020	-0.04	-0.27	0.52	4.01	0.25	0.69
82021	-0.01	-0.26	0.50	11.94	0.11	0.94
82022	0.05	0.69	0.41	4.99	0.20	0.94
82031	0.12	0.91	0.28	1.88	0.84	2.86
82100	0.44	2.94	0.17	1.01	-0.82	-1.14
82101	0.00	-0.01	0.25	2.15	0.53	2.04
82123	0.03	0.21	0.18	1.37	-0.23	-0.59
Secondary Diagnoses						
E88	-0.12	-2.54	—	—	—	—
198	—	—	-0.28	-1.97	-0.76	-1.50
250	—	—	—	—	0.20	1.45
276	—	—	0.15	2.30	—	—
281	—	—	0.34	2.49	—	—
285	—	—	0.10	1.92	—	—
290	-1.24	-11.43	0.22	2.84	-1.33	-3.21
298	-0.89	-5.26	—	—	-1.95	-1.96
310	-1.13	-12.59	—	—	-0.46	-1.93
331	-0.94	-6.22	—	—	—	—
332	—	—	—	—	0.87	4.01
401	0.25	5.18	—	—	0.41	3.37
402	0.25	2.43	0.29	2.85	—	—
414	—	—	0.13	2.71	—	—
415	0.33	1.90	—	—	—	—
427	—	—	0.28	5.28	—	—
428	-0.27	-4.08	0.14	2.23	—	—
437	-0.49	-3.88	—	—	—	—
438	—	—	—	—	0.63	2.78
492	—	—	—	—	-0.84	-1.44
599	—	—	0.11	2.47	—	—
707	-0.40	-3.01	0.25	2.07	—	—
714	—	—	—	—	0.63	1.91
788	0.28	2.55	—	—	—	—
797	-1.12	-5.58	—	—	-1.75	-1.75
812	—	—	0.56	5.26	—	—
813	—	—	—	—	0.61	2.62
820	—	—	0.16	1.82	—	—
997	—	—	—	—	0.34	1.50
Other	-0.01	-0.12	0.03	0.83	0.21	1.87
Interact	-0.01	-0.12	0.03	0.81	-0.07	-0.71

Appendix B

REPLICATING THE ESTIMATION PROCESS

The reliability of the regression model estimation procedure used in Secs. III and IV is examined in this appendix. The data for DRG 14 were divided randomly into two subsets of approximately equal size. The results for the first subset are presented in Sec. III. The results for the second subset are given in this appendix for comparison to judge how much the estimates and t-statistics should be deflated to account for the fact that only the largest estimates from a large pool of candidates were included in the final modeling. Each subset is approximately the same size as the datasets of the remaining DRG categories, so the results for DRG 14 should be indicative of the other DRG categories.

Tables B.1 to B.3 contain the results for the HHA, SNF, and REHAB models. The estimates from each replication are placed side by side to make comparisons easier. As was anticipated, there is a substantial reduction in the magnitudes of the estimates and t-statistics in the second replication for secondary diagnoses identified as having high predictive value. Even with this reduction, most of the secondaries identified as having high predictive value in the first replication retain the same sign and continue to show evidence that they are related to the posthospital variables. There are numerous secondaries with t-statistics larger than 4.0, and their presence cannot be explained by selection effects alone.

A more systematic comparison of the estimates from the two independent replications can be used to identify pairs of estimates that are unusually different. This comparison also provides an opportunity to verify that the standard errors, and normal approximations used with them, are providing t-statistics and confidence intervals with the inferential distributions claimed for them. By construction, the estimates from each half of the data are independent estimates of the same population values. Denote the two independent estimates for the j^{th} secondary diagnosis by β_j^1 and β_j^2. The usual asymptotic theory, on which all of our statistical results are based, predicts that the estimates follow approximately normal distributions about their population values with standard deviations given by their standard errors,

$$\left(\hat{\beta}_j^1 - \beta_j\right) \sim N\left(0, (se_1)^2\right), \quad \left(\hat{\beta}_j^2 - \beta_j\right) \sim N\left(0, (se_2)^2\right).$$

The difference in the two estimates is thus predicted to follow a normal distribution with a variance derived from the standard errors of each estimator,

$$\left(\hat{\beta}_j^1 - \hat{\beta}_j^2\right) \sim N\left(0, (se_1)^2 + (se_2)^2\right).$$

As noted previously, the secondary diagnoses have low correlation with each other, so the parameter estimates for the secondary diagnoses are also approximately uncorrelated. The differences in the estimated parameters for each secondary diagnosis thus provide 44 replications to test the asymptotic approximation. Figures B.1 to B.3 contain quantile plots of the differences between the two parameter estimates for each secondary diagnosis plotted against the expected normal order statistics under the asymptotic approximation for HHA, SNF, and REHAB. Each difference has been divided by its standard error so that these scaled differences should be like a sample of size 44 from the standard normal distribution and fall along the line passing through zero with slope one. Agreement with the behavior predicted by the aysmptotic distributions is very good. The only exception to this statement is for the parameter estimates of secondary diagnoses 427 and 787 when predicting SNF use. All of the estimates for these secondary diagnoses indicate a strong dependence between SNF use and the presence of these secondary diagnoses. The divergence from the asymptotic theory when there is a strong dependence like this is well known in the statistical literature. The normal approximation usually fails because the tails of the sampling distributions of the parameters are heavier than predicted. Distributions skewed towards parameter estimates with large magnitude are also common, for example, Jennings (1986).

Summarizing the results of this appendix, there is evidence of substantial exaggeration in the magnitude of the secondary diagnoses with larger parameter estimates and their corresponding large t-statistics. A crude conservative guide, developed through casual examination of the two sets of estimators, would be to shrink each estimate and t-statistic by about 25 percent. Even with this adjustment, there is still evidence that the secondary diagnoses could be used to provide predictive information about posthospital use not contained in the principal diagnoses. This is consistent with the likelihood ratio chi square statistics presented in Secs. III and IV. Examination of the differences between the two sets of parameter estimates provides evidence to support the use of the standard asymptotic approximations for logistic regression estimators.

Table B.1

REPLICATION OF THE REGRESSION ESTIMATES: HHA

	Full				Reduced			
	First Replication		Second Replication		First Replication		Second Replication	
Code	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
				Primary Diagnoses				
430	-0.42	—	-0.50	—	-0.69	—	-0.59	—
431	-0.09	—	-0.13	—	-0.09	—	-0.11	—
4320	-1.28	—	-0.92	—	-6.08	—	-0.82	—
4321	-0.67	—	-0.65	—	-0.90	—	-0.91	—
4329	-0.35	—	-0.20	—	-0.40	—	-0.13	—
4340	0.10	—	0.08	—	0.10	—	0.09	—
4341	0.13	—	0.04	—	0.13	—	0.06	—
4349	0.13	—	0.09	—	0.13	—	0.07	—
4373	-0.83	—	-0.79	—	-1.27	—	-1.06	—
7843	-0.42	—	-0.56	—	-0.44	—	-0.64	—
				Secondary Diagnoses				
V10	0.04	0.31	0.06	0.52	—	—	—	—
41	-0.01	-0.07	-0.04	-0.42	—	—	—	—
244	0.17	1.26	-0.09	-0.66	—	—	—	—
250	0.21	5.33	0.23	5.88	0.19	4.69	0.21	5.19
276	-0.05	-0.82	-0.06	-0.95	—	—	—	—
278	-0.18	-1.28	0.03	0.20	—	—	—	—
285	0.18	1.45	0.12	0.99	—	—	—	—
290	-0.20	-1.84	0.12	1.11	-0.21	-1.80	—	—
298	-0.02	-0.15	-0.29	-2.17	—	—	-0.32	-2.15
310	-0.09	-1.01	-0.28	-3.03	—	—	-0.31	-3.07
331	0.07	0.75	-0.23	-2.55	—	—	-0.25	-2.54
332	0.05	0.41	0.11	0.85	—	—	—	—
342	0.23	6.49	0.20	5.54	0.23	6.06	0.19	4.88
344	0.32	2.92	0.30	2.91	0.30	2.90	0.29	2.87
345	0.01	0.08	-0.07	-0.48	—	—	—	—
401	0.14	4.21	0.14	4.00	0.13	3.41	0.12	3.18
402	0.18	2.68	0.27	4.15	0.16	2.37	0.24	3.76
410	0.05	0.39	0.18	1.34	—	—	—	—
412	0.06	0.57	-0.12	-1.15	—	—	—	—
413	0.08	0.74	-0.02	-0.23	—	—	—	—
414	-0.05	-1.07	-0.05	-1.18	—	—	—	—
424	-0.08	-0.71	-0.12	-1.14	—	—	—	—
426	0.16	1.67	0.24	2.36	—	—	0.23	2.31
427	-0.05	-1.23	-0.02	-0.58	—	—	—	—
428	0.10	1.76	0.07	1.32	—	—	—	—
429	0.04	0.60	-0.08	-1.20	—	—	—	—
433	-0.14	-1.92	-0.05	-0.69	-0.15	-1.97	—	—
434	0.30	2.58	0.14	1.22	0.28	2.43	—	—
435	-0.14	-1.36	-0.22	-2.25	—	—	-0.26	-2.39
436	0.24	1.97	0.12	0.99	—	—	—	—
437	0.02	0.23	-0.01	-0.20	—	—	—	—
438	0.12	1.33	0.04	0.41	—	—	—	—
440	-0.00	-0.04	0.00	0.00	—	—	—	—
443	0.06	0.41	0.15	1.09	—	—	—	—
486	-0.34	-3.47	-0.09	-0.95	-0.37	-3.40	—	—
496	-0.02	-0.27	0.14	1.95	—	—	0.12	1.73
507	-0.30	-2.37	-0.33	-2.54	-0.33	-2.37	-0.34	-2.42
599	0.12	2.10	0.08	1.48	0.11	2.15	—	—
715	-0.00	-0.05	-0.02	-0.28	—	—	—	—
780	-0.23	-3.72	-0.20	-3.21	-0.26	-3.77	-0.21	-3.17
784	0.04	0.67	0.02	0.44	—	—	—	—
787	-0.11	-0.86	0.07	0.56	—	—	—	—
788	0.39	2.91	0.11	0.86	0.35	2.82	—	—
Other	-0.02	-0.64	-0.02	-0.78	-0.04	-0.90	-0.06	-1.41
Interact	—	—	—	—	0.02	0.46	0.03	0.61

Table B.2

REPLICATION OF THE REGRESSION ESTIMATES: SNF

	Full				Reduced			
	First Replication		Second Replication		First Replication		Second Replication	
Code	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
				Primary Diagnoses				
4340	0.07	—	0.23	—	0.09	—	0.14	—
4341	-0.26	—	0.02	—	-0.27	—	0.01	—
4349	0.13	—	0.20	—	0.13	—	0.19	—
4373	-0.44	—	-0.54	—	-0.82	—	-1.16	—
7843	0.02	—	-0.49	—	0.10	—	-0.79	—
				Secondary Diagnoses				
V10	0.23	1.57	0.20	1.33	—	—	—	—
V45	0.09	0.65	-0.03	-0.21	—	—	—	—
41	0.73	6.64	0.46	4.08	0.50	5.54	0.32	3.43
244	0.51	2.99	0.06	0.35	—	—	—	—
250	0.13	2.59	0.09	1.80	—	—	—	—
276	0.45	5.87	0.29	3.76	0.39	5.68	0.27	3.83
278	-0.19	-1.06	0.05	0.28	—	—	—	—
285	0.17	1.07	0.41	2.60	—	—	—	—
290	0.37	2.72	0.48	3.57	—	—	0.45	3.80
298	0.12	0.67	0.17	0.97	—	—	—	—
310	0.27	2.34	0.20	1.71	—	—	—	—
331	-0.01	-0.05	0.19	1.66	—	—	—	—
332	0.32	1.88	0.47	2.83	—	—	—	—
342	0.63	13.83	0.68	14.99	0.50	11.31	0.58	13.05
344	0.44	3.19	0.23	1.72	—	—	—	—
345	-0.08	-0.49	0.09	0.52	—	—	—	—
401	-0.03	-0.72	-0.05	-1.16	—	—	—	—
402	0.15	1.85	-0.08	-0.96	—	—	—	—
410	0.11	0.65	-0.02	-0.11	—	—	—	—
412	-0.34	-2.48	-0.02	-0.17	-0.47	-2.89	—	—
413	-0.23	-1.74	-0.21	-1.65	—	—	—	—
414	0.05	0.96	0.04	0.74	—	—	—	—
424	0.09	0.66	-0.02	-0.17	—	—	—	—
426	-0.14	-1.11	0.02	0.18	—	—	—	—
427	0.44	8.20	0.18	3.39	0.37	7.54	0.15	2.99
428	0.38	5.40	0.37	5.32	0.32	5.00	0.32	5.16
429	0.12	1.44	0.11	1.24	—	—	—	—
433	-0.35	-3.77	-0.40	-4.15	-0.55	-4.62	-0.55	-4.61
434	0.06	0.44	0.34	2.32	—	—	—	—
435	-0.50	-3.90	-0.41	-3.27	-0.82	-4.44	-0.63	-3.76
436	0.48	3.04	0.14	0.90	—	—	—	—
437	0.00	-0.04	-0.11	-1.32	—	—	—	—
438	-0.04	-0.38	0.24	2.18	—	—	—	—
440	0.04	0.38	0.01	0.08	—	—	—	—
443	-0.10	-0.58	0.22	1.29	—	—	—	—
486	0.84	6.80	1.16	9.53	0.64	6.23	0.87	9.10
496	0.11	1.27	0.08	0.95	—	—	—	—
507	1.41	8.84	1.36	8.32	0.96	7.92	0.96	7.66
599	0.48	6.92	0.71	10.22	0.39	6.17	0.59	9.92
715	0.11	1.03	0.16	1.53	—	—	—	—
780	0.22	2.79	0.30	3.87	—	—	0.26	3.57
784	0.36	5.30	0.26	3.87	0.28	4.66	0.22	3.67
787	1.15	7.11	0.48	3.08	0.82	6.56	—	—
788	0.51	3.00	0.43	2.66	—	—	—	—
Other	0.10	2.73	0.05	1.28	0.09	1.60	0.06	1.19
Interact	—	—	—	—	0.11	2.20	0.06	1.19

Table B.3

REPLICATION OF THE REGRESSION ESTIMATES: REHAB

	Full				Reduced			
	First Replication		Second Replication		First Replication		Second Replication	
Code	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.	Est.	t-stat.
Primary Diagnoses								
4340	0.13	—	0.12	—	0.15	—	0.12	—
4341	0.39	—	0.28	—	0.45	—	0.41	—
4349	0.46	—	0.47	—	0.47	—	0.45	—
4373	-0.62	—	-0.26	—	-0.78	—	-0.55	—
7843	-0.22	—	0.21	—	-0.40	—	0.24	—
Secondary Diagnoses								
V10	0.02	0.10	-0.31	-1.47	—	—	—	—
V45	0.29	1.51	0.32	1.71	—	—	—	—
41	-0.04	-0.29	-0.29	-1.87	—	—	—	—
244	0.50	2.13	-0.24	-1.01	—	—	—	—
250	0.18	2.64	0.14	2.06	0.11	1.64	—	—
276	-0.30	-2.88	-0.26	-2.49	-0.46	-3.54	-0.39	-3.03
278	-0.31	-1.25	0.58	2.34	—	—	—	—
285	-0.45	-2.07	-0.35	-1.57	—	—	—	—
290	-0.59	-3.20	-0.41	-2.21	-1.36	-3.78	—	—
298	-0.42	-1.79	-0.51	-2.16	—	—	—	—
310	-0.49	-3.16	-0.59	-3.71	-0.98	-3.82	-1.35	-4.39
331	-0.30	-1.81	-0.48	-3.01	—	—	-0.90	-3.72
332	-0.05	-0.24	-0.38	-1.66	—	—	—	—
342	1.29	20.70	1.23	19.48	0.94	16.33	0.90	15.05
344	-0.05	-0.25	0.58	3.18	—	—	0.45	2.80
345	-0.39	-1.65	0.04	0.15	—	—	—	—
401	0.39	6.63	0.48	8.04	0.30	5.05	0.37	6.19
402	0.27	2.33	0.17	1.48	0.20	1.86	—	—
410	0.96	4.09	0.46	1.90	0.67	3.75	—	—
412	0.41	2.22	0.36	1.96	—	—	—	—
413	0.23	1.28	0.23	1.30	—	—	—	—
414	-0.17	-2.21	-0.10	-1.38	-0.16	-2.09	—	—
424	0.19	1.03	0.33	1.71	—	—	—	—
426	0.09	0.50	0.29	1.61	—	—	—	—
427	0.14	1.84	0.22	3.01	—	—	0.18	2.58
428	-0.11	-1.16	-0.07	-0.77	—	—	—	—
429	-0.29	-2.44	0.01	0.09	-0.37	-2.76	—	—
433	0.22	1.73	0.13	1.00	—	—	—	—
434	0.15	0.77	0.31	1.52	—	—	—	—
435	-0.44	-2.53	-0.64	-3.69	-0.70	-2.81	-1.31	-4.06
436	0.00	0.02	-0.19	-0.88	—	—	—	—
437	-0.15	-1.26	-0.24	-2.04	—	—	—	—
438	-0.08	-0.50	-0.39	-2.55	—	—	-0.40	-2.40
440	-0.17	-1.34	-0.11	-0.88	—	—	—	—
443	0.22	0.97	0.63	2.63	—	—	0.51	2.53
486	-0.32	-1.90	-0.11	-0.63	—	—	—	—
496	-0.01	-0.06	0.09	0.72	—	—	—	—
507	0.07	0.31	0.05	0.20	—	—	—	—
599	0.05	0.49	0.30	3.14	—	—	0.16	2.04
715	0.03	0.22	0.00	0.02	—	—	—	—
780	-0.35	-3.33	-0.47	-4.42	-0.53	-3.91	-0.76	-5.04
784	0.53	5.68	0.53	5.74	0.33	4.44	0.35	4.60
787	-0.02	-0.11	0.50	2.30	—	—	—	—
788	0.26	1.12	-0.18	-0.80	—	—	—	—
Other	0.12	2.33	-0.02	-0.42	-0.02	-0.25	-0.09	-1.19
Interact	—	—	—	—	-0.18	2.24	0.24	3.12

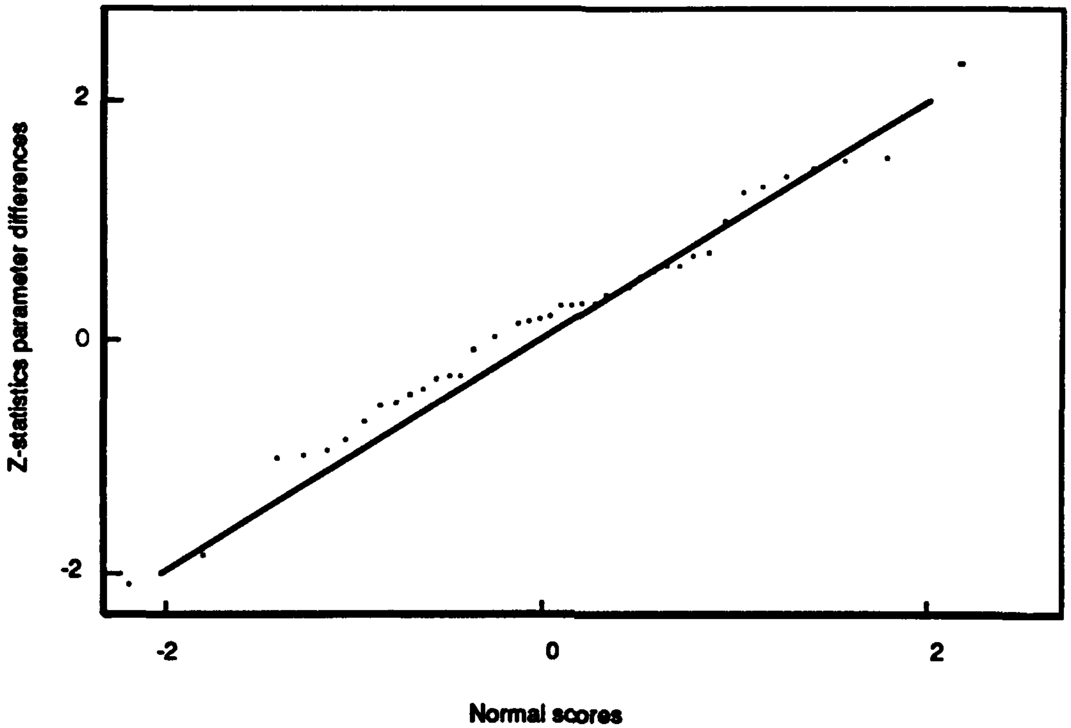

Fig. B.1—Quantile plot for HHA parameters

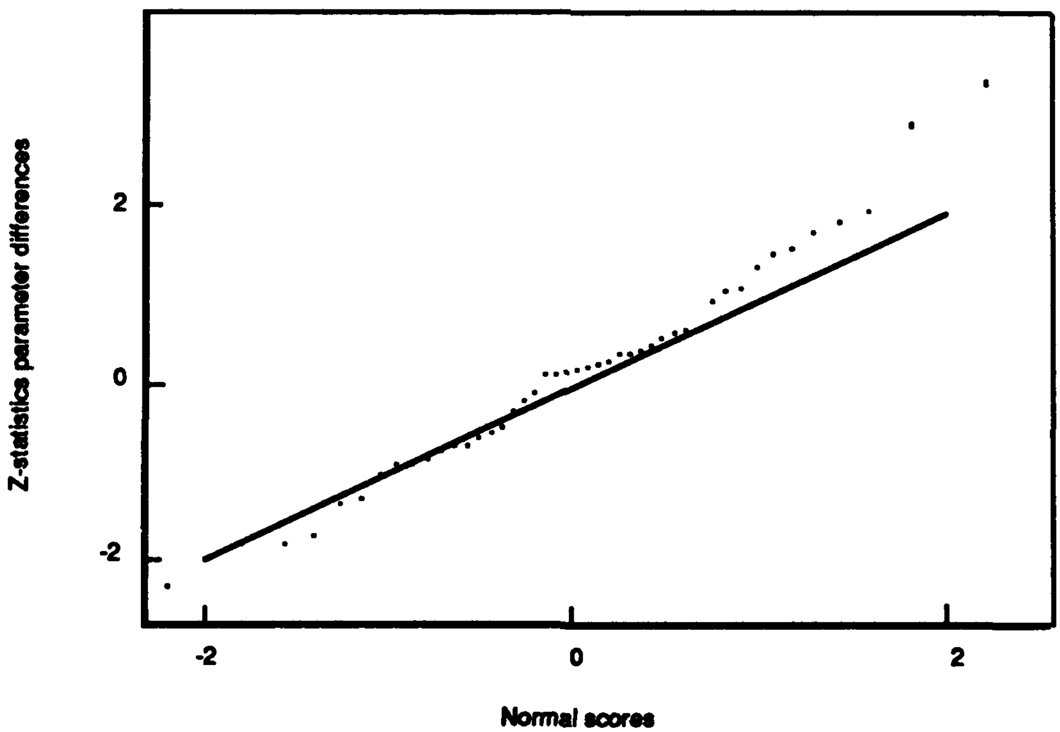

Fig. B.2— Quantile plot for SNF parameters

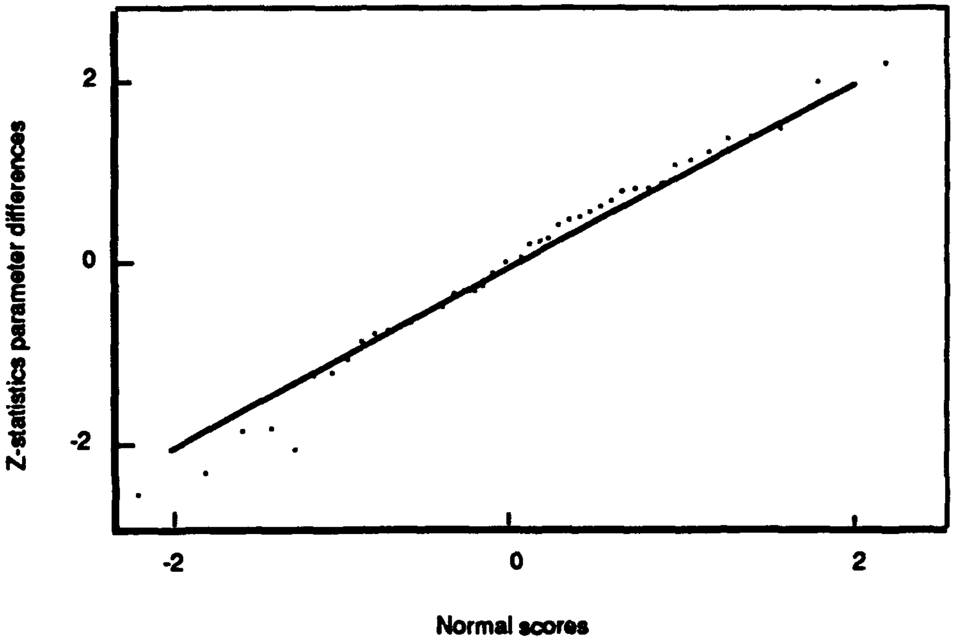

Fig. B.3—Quantile plot for REHAB parameters

REFERENCES

Bishop, Y. M., S. E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice*, The MIT Press, Cambridge, Massachusetts, 1975.

Haggstrom, G., "Logistic Regression and Discriminant Analysis by Ordinary Least Squares," *J. Busi. and Econ. Statist.*, Vol. 3, pp. 229–238, 1983.

Jennings, D. E., "Judging Inference Adequacy in Logistic Regression," *J. Amer. Statist. Assoc.*, Vol. 81, pp. 471–476, 1986.

Neu, C. R., and S. C. Harrison, *Prospective Payment for Medicare Posthospital Services: Some Empirical Considerations*, The RAND Corporation, R-3435-HCFA, December 1986.

Neu, C. R., S. C. Harrison, and J. Z. Heilbrun, *Medicare Patients and Postacute Care: Who Goes Where?* The RAND Corporation, R-3780-MN, November 1989.